

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉTUDES DE BIOÉQUIVALENCE

MÉMOIRE  
PRÉSENTÉ  
COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN MATHÉMATIQUES

PAR  
ANGELA MIRZAC

DÉCEMBRE 2008

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

Je tiens tout d'abord remercier ma directrice de mémoire, Sorana Froda, professeure, département de mathématiques de l'UQAM, pour sa disponibilité, son encadrement et son encouragement. J'ai toujours bien apprécié son côté professionnel et celui humain également. Merci.

Sincères remerciements à Is-Serge Alalouf, mon premier professeur du département de mathématiques de l'UQAM, qui a su me soutenir et m'encourager tout au long de mes études.

Je remercie également Bertrand Fournier, analyste, Service de Consultation en Analyse de Données (SCAD), UQAM, de m'avoir offert la possibilité de réaliser des simulations dans le laboratoire du SCAD.

Je n'oublierais Gisèle Legault, analyste de l'informatique, Laboratoire des cycles supérieurs en mathématiques, pour son soutien informatique et Manon Gauthier, assistante gestion programme cycles supérieurs, pour ses conseils et son aide à la distance.

Un gros merci à Véronique Pagé, chef du département de la statistique, MDS Pharma Service, de m'avoir permise de me libérer pour deux mois du travail afin que je puisse me consacrer temps plein aux recherches et à la finalisation de mon mémoire.

Mes remerciements particuliers à mon mari, à ma fille et, bien sur, à mes parents qui ont toujours su à me soutenir et à m'encourager malgré la distance qui nous sépare.

## TABLE DES MATIÈRES

LISTE DES FIGURES.....	V
LISTE DES TABLEAUX.....	VI
RÉSUMÉ .....	VIII
INTRODUCTION .....	1
CHAPITRE I	
BIOÉQUIVALENCE (QUESTION PRATIQUE).....	4
1.1 Paramètres pharmacocinétiques.....	7
1.2 Mesures de biodisponibilité.....	14
1.3 Exemple de calcul et estimation des paramètres pharmacocinétiques.....	17
CHAPITRE II	
TESTS EN BIOÉQUIVALENCE.....	20
2.1 Généralités sur les tests d'hypothèses.....	21
2.2 Tests classiques de comparaison des moyennes .....	22
2.2.1 Échantillons indépendants .....	22
2.2.2 Observations appariées.....	30
2.3 Motivation pratique des tests utilisés en bioéquivalence.....	34
2.4 Tests d'équivalence.....	36
2.4.1 Test unilatéral (non infériorité).....	37
2.5 Principe d'intervalle de confiance .....	45
CHAPITRE III	
PLANS D'EXPÉRIENCE EN BIOÉQUIVALENCE.....	48
3.1 Plan d'expérience parallèle.....	49
3.2 Plan croisé $g \times p$ .....	52
3.2.1 Exemples de plans croisés .....	54
3.2.2 Modèle linéaire général .....	58
3.3 Inférence statistique pour le plan croisé $2 \times 2$ (tests classiques).....	60
3.3.1 L'effet résidu .....	64

3.3.2 L'effet direct du traitement.....	68
3.3.3 L'effet période.....	72
3.4 L'analyse de variance pour le plan croisé 2 x 2.....	73
3.5 Tests de bioéquivalence.....	80
CHAPITRE IV	
ÉTUDE DE SIMULATION.....	90
4.1 Description de l'expérience et de ses paramètres.....	90
4.2 Simulation des erreurs expérimentales.....	93
4.3 Tests effectués.....	97
4.4 Résultats.....	103
CONCLUSION.....	110
APPENDICE A	
RÈGLE TRAPÉZOÏDALE.....	112
APPENDICE B	
MACRO SAS.....	114
BIBLIOGRAPHIE.....	117

## LISTE DES FIGURES

Figure		page
Figure 1.1	«Égalité» des moyennes et des variances .....	6
Figure 1.2	«Égalité» des moyennes mais pas des variances .....	6
Figure 1.3	«Égalité» des variances mais pas des moyennes .....	7
Figure 1.4	Concentrations plasmatiques après l'administration d'un traitement .....	9
Figure 1.5	Concentrations plasmatiques après l'administration intraveineuse d'un traitement.....	10
Figure 1.6	Concentrations plasmatiques (transformées en $\log_{10}$ près l'administration intraveineuse d'un traitement (selon la formule (1.6)) .....	13
Figure 3.1	Loi log normale .....	82
Figure 4.1	Histogramme de loi de mélange composée de 70% loi normale et de 30% loi double-exponentielle et densité de loi normale .....	101
Figure 4.2	Histogramme de loi de mélange composée de 60% loi normale et de 40% loi double-exponentielle et densité de loi normale .....	102
Figure 4.3	Loi double-exponentielle .....	102

## LISTE DES TABLEAUX

Tableau	page
Tableau 1.1	Concentrations sanguines de primidone ( $mg/mL$ ) obtenues suite à l'administration du traitement à un sujet. .... 18
Tableau 2.1	Toutes les permutations possibles de 7 résultats obtenus suite à l'administration du traitement $A$ à $n = 4$ sujets et du traitement $B$ à $m = 3$ sujets. .... 28
Tableau 2.2	Concentrations maximales de norgestrel ( $ng/ml$ ) dans le sérum sanguin des 28 femmes pour la consommation d'olestra, $X_i$ , et la consommation des repas contenant du triglycéride ordinaire, $Y_i$ ... 43
Tableau 3.1	Exemple de plan parallèle..... 50
Tableau 3.2	Analyse de Variance pour un plan d'expérience parallèle ..... 51
Tableau 3.3	Plan croisé $2 \times 2$ ..... 54
Tableau 3.4	Séquences du plan croisé $2 \times 3$ ..... 56
Tableau 3.5	Séquences du plan croisé $2 \times 4$ ..... 56
Tableau 3.6	Séquences du plan croisé $3 \times 3$ ..... 57
Tableau 3.7	Séquences du plan croisé $4 \times 4$ ..... 58
Tableau 3.8	Combinaisons période-séquence du modèle linéaire du plan croisé $2 \times 2$ ..... 61
Tableau 3.9	Les effets fixes du plan croisé $2 \times 2$ sans effet séquence..... 62
Tableau 3.10	Les effets fixes du plan croisé $2 \times 2$ avec effet séquence ..... 63
Tableau 3.11	Les effets aléatoires du plan croisé $2 \times 2$ ..... 63

Tableau 3.12	Les espérances mathématiques des variables $u_{ik}$ .....	65
Tableau 3.13	Les espérances mathématiques des variables $d_{ik}$ .....	69
Tableau 3.14	Analyse de la Variance .....	79
Tableau 3.15	Paramètres importants de la loi normale et de la loi log-normale	83
Tableau 4.1	L'espérance mathématique de la différence entre l'observation de la période 1 et l'observation de la période 2 du sujet $i$ ( $i = 1, 2, \dots, n$ ) dans la séquence 1 ou 2. ....	98
Tableau 4.2	Nombre total des permutations pour différentes tailles échantillonales. ....	100
Tableau 4.3	Comparaison des valeurs-p (pour un seuil de signification $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi normale ; $\sigma_B^2 = 0.015$ ; $\sigma_W^2 = 0.01$ (500 simulations). ....	103
Tableau 4.4	Comparaison des valeurs-p (pour un seuil de signification $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi normale ; $\sigma_B^2 = 0.015$ ; $\sigma_W^2 = 0.01$ (1000 simulations). ....	104
Tableau 4.5	Comparaison des valeurs-p (pour un seuil de signification $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi de mélange de loi normale (70%) et de loi double-exponentielle (30%) ; $\sigma_B^2 = 0.045$ ; $\sigma_W^2 = 0.03$ .....	105
Tableau 4.6	Comparaison des valeurs-p (pour un seuil de signification $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi de mélange de loi normale (60%) et de loi double-exponentielle (40%) ; $\sigma_B^2 = 0.045$ ; $\sigma_W^2 = 0.03$ .....	106
Tableau 4.7	Comparaison des valeurs-p (pour un seuil de signification $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi double-exponentielle; $\sigma_B^2 = 0.045$ ; $\sigma_W^2 = 0.03$ . ....	107

## RÉSUMÉ

Dans la pratique clinique, on fait souvent appel aux études de bioéquivalence afin de comparer deux médicaments. Un premier objectif de ce mémoire est d'introduire la notion pharmacocinétique et statistique de ce type d'étude, ainsi que de définir et de présenter l'application des tests d'équivalence à travers les études de bioéquivalence.

Pour réaliser l'inférence statistique des études de bioéquivalence, en pratique on utilise le test  $t$  de Student. Le second but de ce mémoire est l'utilisation du test de permutations dans le cas où les hypothèses du test paramétrique ne sont pas satisfaites. À travers une étude de simulation des données de plusieurs lois on compare la performance du test de permutations avec le test  $t$  de Student et le test de Wilcoxon pour deux échantillons. Ce dernier est le test non paramétrique utilisé dans la pratique courante des études de bioéquivalence.

Mots clés : bioéquivalence, tests d'hypothèses, plan d'expérience croisé, test de permutation.

## INTRODUCTION

De point de vue clinique les études de bioéquivalence consistent à montrer que deux ou plusieurs médicaments qui contiennent le même ingrédient actif donnent le même effet thérapeutique. Le plus souvent, ce type d'étude se fait pour deux médicaments où le patent d'un médicament déjà sur le marché est en cours d'expiration et qu'un médicament générique est créé pour le remplacer.

De point de vue statistique les études de bioéquivalence consistent à démontrer, par le biais de l'inférence statistique, l'équivalence en termes de mesures de **biodisponibilité** (voir chapitre 1) du nouveau médicament et du médicament standard. Dans ce mémoire, on appelle alors par traitement **Référence** le médicament standard et par traitement **Test** le nouveau médicament.

Généralement, pour comparer deux moyennes  $\mu_T$  et  $\mu_R$  on met  $\theta = \mu_T - \mu_R$  et on définit un test de deux hypothèses concurrentes de la façon suivante :

$$H_0: \theta = \theta_0$$

contre l'une des hypothèses

$$H_a: \theta \neq \theta_0, H_a: \theta < \theta_0 \text{ ou bien } H_a: \theta > \theta_0.$$

Dans le cadre d'une étude de bioéquivalence on inverse l'hypothèse nulle et l'hypothèse alternative. De plus, contrairement à un test ordinaire, l'hypothèse alternative d'une étude de bioéquivalence est définie par un intervalle d'équivalence comme dans l'exemple suivant :

$$H_0: \theta \leq \theta_0 - \varepsilon_1 \text{ ou } \theta \geq \theta_0 + \varepsilon_2,$$

d'où l'alternative

$$H_a: \theta_0 - \varepsilon_1 < \theta < \theta_0 + \varepsilon_2,$$

avec  $\varepsilon_1$  et  $\varepsilon_2$  connues. Ainsi, on prend l'**inéquivalence** des traitements Test et Référence comme hypothèse nulle et l'**équivalence** des traitements comme hypothèse alternative.

En pratique, les valeurs  $\varepsilon_1$  et  $\varepsilon_2$  sont établies d'avance par des agences qui contrôlent la mise sur le marché des nouveaux médicaments. Aux États-Unis, l'agence FDA (Food and Drug Administration), propose ce qu'on appelle la **règle  $\pm 20$**  comme critère de base pour les études de bioéquivalence. D'après cette règle, la bioéquivalence est conclue, si la moyenne du traitement Test n'excède pas les limites de  $\pm 20\%$  de la moyenne du traitement Référence. En d'autres termes, le rapport des moyennes  $\mu_T/\mu_R$  ne doit pas dépasser les limites  $[80\%, 120\%]$ . Notons que, dans la pratique courante, les valeurs des mesures de biodisponibilité qui servent à évaluer la bioéquivalence des traitements sont souvent transformées en logarithme naturel dans le but d'obtenir des données qui suivent des lois normales. Dans ce contexte, ce qu'on appelle **intervalle de bioéquivalence** revient à  $[80\%, 125\%]$  (voir chapitre 3, section 3.5).

Ce mémoire vise à définir brièvement les études de bioéquivalence en termes pharmacocinétiques, à présenter plus en détail les tests d'équivalence, dont la bioéquivalence est un cas pratique, ainsi que leur motivation statistique. On se concentre aussi sur les plans d'expérience croisés qui sont le plus souvent utilisés pour ce type d'étude, ainsi qu'à l'application et à l'analyse des résultats des tests statistiques courants dans le cadre des études de bioéquivalence. De plus, on propose d'utiliser des tests de permutations dans ce contexte et on compare leur performance avec les tests habituels de la pratique pharmacocinétique (test de t de Student et test de Wilcoxon pour deux échantillons). Cette comparaison se fait à travers une étude de simulation.

Donc, une brève introduction à la pharmacocinétique, aux principaux paramètres pharmacocinétiques, ainsi qu'aux mesures de biodisponibilité est donnée dans le chapitre 1. À la fin de ce chapitre on présente un exemple de calcul et d'estimation de quelques paramètres pharmacocinétiques.

Le chapitre 2 débute avec un rappel des notions générales sur les tests d'hypothèses et l'inférence statistique, ainsi qu'avec une brève description de certains tests classiques pour des échantillons indépendants et appariés. On parle plus en détails du test non paramétrique de permutation qui est utilisé ultérieurement dans nos simulations (chapitre 4). On décrit la motivation pratique des tests de bioéquivalence. On introduit la notion de test d'équivalence ainsi que le principe d'intervalle de confiance dans le cadre des études d'équivalence.

Dans le chapitre 3 on motive le choix du plan d'expérience dans le cadre des études de bioéquivalence et on s'arrête plus en détails sur le plan croisé  $2 \times 2$ , le plan le plus utilisé pour ce type d'études. On finit ce chapitre avec une section où on se concentre sur les tests de bioéquivalence, en particulier sur le principe de Schuirmann (1981, 1987) des deux tests unilatéraux d'équivalence.

Finalement, dans le chapitre 4 on décrit étape par étape la méthodologie appliquée pour simuler et analyser des données. On décrit les résultats obtenus par le test de permutations et on les compare à ceux obtenus avec le test t de Student et le test de Wilcoxon pour deux échantillons.

## CHAPITRE I

### BIOÉQUIVALENCE (QUESTION PRATIQUE)

**L'objectif d'une étude de bioéquivalence** est de comparer un nouveau traitement (médicament), appelé Test, abrégé  $T$ , à un traitement (médicament) standard, appelé Référence, abrégé  $R$ , et d'estimer leur **équivalence** en termes de mesures de biodisponibilité. Ces mesures sont : le taux et l'étendue de l'absorption de l'ingrédient actif du traitement dans l'organisme. Généralement, le traitement référence représente le produit de l'innovateur (produit ancien), tandis que le traitement test est un produit générique ou une nouvelle formulation du traitement référence. La **biodisponibilité** d'un médicament est le terme utilisé pour indiquer la quantité totale de médicament et la vitesse avec laquelle cette quantité atteint la circulation sanguine et est disponible au site d'action, après l'administration du médicament. Deux produits pharmaceutiques déclarés bioéquivalents sont supposés aussi équivalents de point de vue thérapeutique. Cela veut dire que ces produits contiennent la même quantité du même ingrédient actif du médicament et ils donnent le même effet thérapeutique.

Le concept statistique de bioéquivalence consiste à comparer les lois marginales des mesures de biodisponibilité des deux traitements.

Dans le contexte de la bioéquivalence, il existe trois types d'études :

- la bioéquivalence moyenne;
- la bioéquivalence de la population;
- la bioéquivalence individuelle.

Pour une étude de **bioéquivalence moyenne** (*average bioequivalence*), on teste «l'égalité» (dans un sens qui sera précisé dans le chapitre 2) qu'on appelle «l'équivalence» des moments d'ordre un des lois marginales de la biodisponibilité du traitement Test et du traitement Référence, respectivement. Pour une étude de **bioéquivalence de population** (*population bioequivalence*), en plus des premiers moments, on teste également «l'égalité» des moments d'ordre deux. Donc, comparativement à la bioéquivalence moyenne, la bioéquivalence de population porte sur la comparaison de la variabilité de la biodisponibilité des deux traitements et pas seulement sur les moyennes. Si une mesure de biodisponibilité du traitement Test et du traitement Référence suit une loi normale, la comparaison se fait entre leurs moyennes et leurs variances, où, si  $M_X(t)$  dénote la fonction génératrice des moments,

$$E[X] = M'(0) = \mu,$$

$$E[X^2] = M''(0) = \sigma^2 + \mu^2,$$

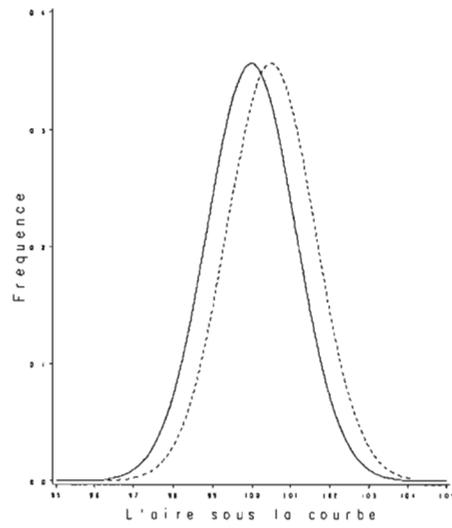
et donc

$$\text{Var}(X) = M''(0) - (M'(0))^2 = \sigma^2.$$

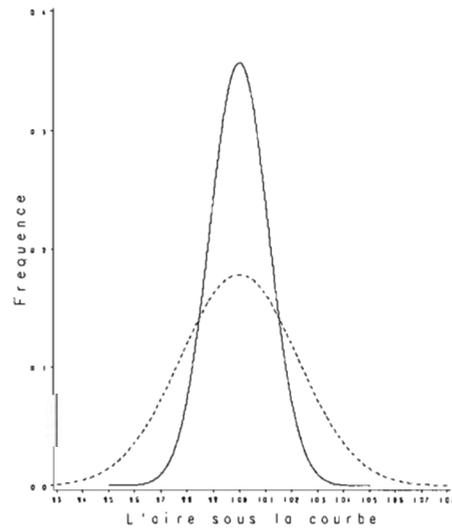
Quelques exemples d'études d'égalité des moyennes et des variances des biodisponibilités des traitements Test et Référence, respectivement, sont présentées dans les figures 1.1-1.3.

Finalement, la **bioéquivalence individuelle** compare le premier et le deuxième moment des lois marginales de la biodisponibilité des deux traitements pour le même sujet. Ces valeurs sont obtenues suite à des administrations répétées du traitement Test et du traitement Référence.

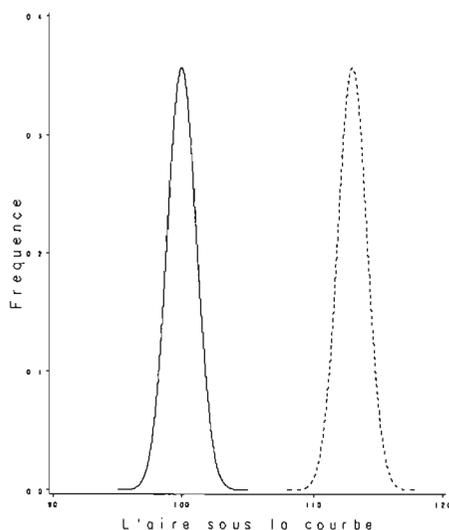
D'après les critères de bioéquivalence de plusieurs agences qui contrôlent la mise en marché des nouveaux traitements, il est suffisant de montrer «l'égalité» des premiers deux moments des biodisponibilités des deux traitements pour que ces traitements soient considérés bioéquivalents.



**Figure 1.1** «Égalité» des moyennes et des variances



**Figure 1.2** «Égalité» des moyennes mais pas des variances



**Figure 1.3** «Égalité» des variances mais pas des moyennes

Dans ce mémoire on traite la bioéquivalence moyenne, et on présente les analyses statistiques usuelles dans ce contexte.

Dans les sections suivantes on introduit brièvement la notion de **pharmacocinétique** et des **paramètres pharmacocinétiques**, et on définit les mesures de biodisponibilité et leurs méthodes de calcul.

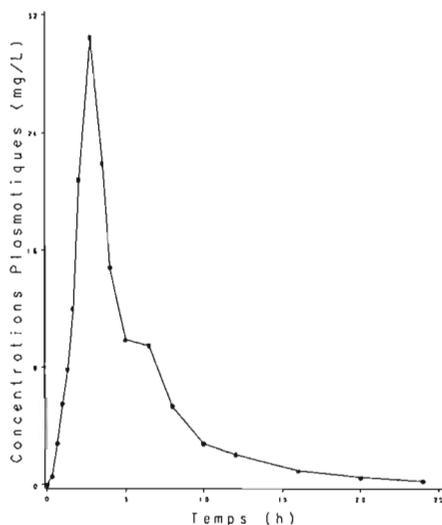
### 1.1 Paramètres pharmacocinétiques

La pharmacocinétique étudie l'absorption, la distribution, le métabolisme et l'élimination d'un médicament de l'organisme. Les mesures de biodisponibilité sont des paramètres pharmacocinétiques. L'estimation des paramètres pharmacocinétiques s'effectue à partir des données expérimentales obtenues chez le sujet suite à l'administration du médicament. Il s'agit, le plus souvent, de l'évolution des concentrations plasmatiques (sanguines) du médicament au cours du temps, mais les données urinaires peuvent également

être utilisées. Après l'administration d'une dose de médicament, ces concentrations plasmatiques (sanguines) varient dû à l'ensemble des transferts entre le milieu extérieur, le sang et les tissus. Leurs valeurs dépendent donc de la vitesse des transferts propres à chacune des étapes pharmacocinétiques.

Les paramètres pharmacocinétiques peuvent être calculés selon une approche **compartimentale** ou **non-compartimentale**. Avec l'approche non-compartimentale, les paramètres pharmacocinétiques sont déduits directement de la courbe obtenue à partir des concentrations plasmatiques (sanguines) observées à des moments de temps fixés d'avance (figure 1.4). Avec l'approche compartimentale, la variation des concentrations est expliquée par la notion de modèle pharmacocinétique compartimental, où l'organisme est représenté par un ou plusieurs compartiments. Le terme **compartiment** désigne le volume fictif dans lequel le médicament se distribue, comme par exemple le volume du sang.

- Si on considère que l'organisme se comporte comme un ensemble homogène, la cinétique (la vitesse des transferts) du médicament est représentée par un modèle **mono-compartimental**.
- Si on considère que l'organisme se comporte comme plusieurs ensembles homogènes, la cinétique du médicament est représentée par un modèle **bi-** ou **pluri-compartimental** avec un compartiment central et un ou plusieurs compartiments périphériques.

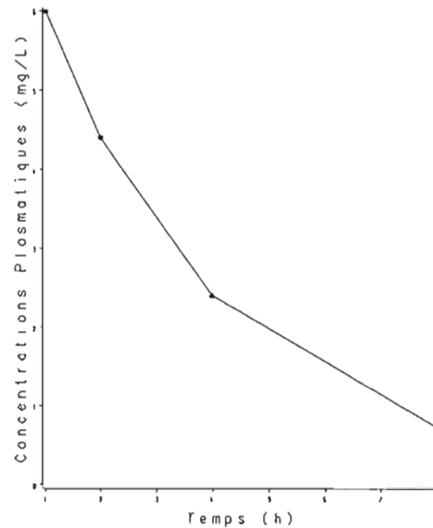


**Figure 1.4** Concentrations plasmatiques après l'administration d'un traitement

Avec cette notion de modèle, l'évolution des concentrations dans le plasma (sang) au cours du temps peut être décrite par des équations mathématiques simples. La résolution de ces équations permet de calculer les paramètres pharmacocinétiques.

Dans ce chapitre, on traite brièvement l'approche mono-compartmentale et non-compartmentale.

Après l'administration d'une dose par voie extravasculaire (buccale, intramusculaire, etc.), la concentration n'augmente pas instantanément à la circulation générale du sang, car le médicament doit franchir des barrières biologiques avant d'y arriver. Cependant, par voie intraveineuse, le médicament est absorbé directement dans le sang (figure 1.5).



**Figure 1.5** Concentrations plasmatiques après l'administration intraveineuse d'un traitement

En ce qui suit on analysera, donc, en lignes générales les transfères du médicament dans l'organisme en considérant les deux cas suivantes d'administration du médicament :

1. par voie intraveineuse;
2. par voie extravasculaire.

### 1. Administration intraveineuse du médicament

L'évolution des concentrations au cours du temps est la résultante de l'entrée du médicament dans l'organisme et de son élimination. Tel que mentionné, par administration intraveineuse le médicament est absorbé directement dans le sang, donc, on n'analyse que l'absorption maximale et la phase de décroissance des concentrations dans ce cas. Dans ce contexte, le taux d'élimination du médicament de l'organisme peut être décrit par l'équation déterministe :

$$\frac{dG}{dt} = -k_e G, \quad (1.1)$$

où  $G(t)$  représente la quantité du médicament dans l'organisme au moment  $t$  après l'administration et  $k_e$  est la **constante du taux d'élimination** qui représente la fraction de médicament éliminée par l'organisme par unité de temps. Pour décrire les transferts du médicament dans l'organisme après son administration, on intègre l'équation (1.1). Comme résultat on obtient

$$G = G_0 e^{-k_e t}, \quad (1.2)$$

une fonction exponentielle décroissante.

En transformant les deux parties de l'équation (1.2) en logarithme naturel, on obtient

$$\ln(G) = \ln(G_0) - k_e t. \quad (1.3)$$

Étant donné que  $\ln(a) = 2.303 \log_{10}(a)$ , on a aussi

$$\log_{10}(G) = \log_{10}(G_0) - \frac{k_e t}{2.303}. \quad (1.4)$$

La relation entre la quantité de médicament dans l'organisme ( $G$ ) et la concentration du médicament dans le plasma ou sang ( $C$ ) est habituellement décrite par la formule

$$G = VC, \quad (1.5)$$

où  $V$  est une constante et représente le **volume apparent** de distribution du médicament dans l'organisme. On dit apparent car  $V$  n'a pas de réalité physiologique, c'est-à-dire il ne fait pas de référence au volume réel. Ainsi, la valeur du  $V$  pour un sujet, dont le poids est de 70 kg, peut être de quelques centaines de litres. Ce paramètre correspond au volume fictif dans lequel le médicament devrait être réparti pour être dans la même concentration que dans le plasma (sang). Ainsi, la constante  $V$  traduit l'intensité de la distribution du médicament dans l'organisme.

En substituant (1.5) dans (1.4) on obtient une transformation de la relation quantité-temps en relation concentration-temps équivalente à  $C = C_0 e^{-k_e t}$ , c'est-à-dire :

$$\log_{10}(C) = \log_{10}(C_0) - \frac{k_e t}{2.303}. \quad (1.6)$$

La valeur  $C_0$  représente la concentration du médicament dans le plasma (sang) immédiatement après administration. D'après l'équation obtenue, on constate que la courbe  $\log_{10}(C)$  versus temps  $t$  est linéaire avec une pente définie par la formule suivante (voir aussi figure 1.6) :

$$\lambda = \frac{k_e t}{2.303}.$$

La valeur de la constante d'élimination  $k_e$  peut s'obtenir à partir de la valeur de la pente,  $\lambda$ , ou à partir de l'équation suivante :

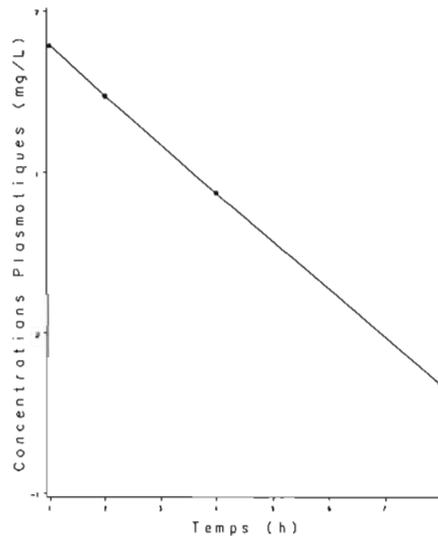
$$k_e = \frac{0.693}{t_{1/2}}, \quad (1.7)$$

où le paramètre pharmacocinétique  $t_{1/2}$  représente la **demi-vie** du médicament. Sa valeur correspond au temps nécessaire pour que, après l'administration d'un médicament, sa concentration diminue de moitié. La demi-vie est exprimée en unités de temps et peut varier de quelques minutes à plusieurs semaines selon le médicament. La fraction de médicament éliminée dans le temps dépend donc de sa demi-vie. On considère que presque la totalité de médicament (97%) est éliminée au bout de 5 demi-vies.

## 2. Administration extravasculaire du médicament

L'administration d'un médicament par voie intraveineuse est réservée aux situations où un effet rapide est recherché ou lorsque les médicaments ne peuvent pas être administrés par voie extravasculaire, car ils sont peu ou mal absorbés. L'étape d'absorption existe pour toutes les voies d'administration extravasculaire (buccale, cutanée, intramusculaire etc). Elle peut

être accompagnée d'une perte en médicament, correspondant à une fraction non absorbée qui n'atteindra pas la circulation générale.



**Figure 1.6** Concentrations plasmatiques (transformées en  $\log_{10}$  près l'administration intraveineuse d'un traitement (selon la formule (1.6))

Contrairement à l'administration intraveineuse, les processus d'absorption et d'élimination coexistent et l'aspect de la courbe varie avec les durées respectives de chacune de ces phases (figure 1.4):

- Phase d'augmentation des concentrations: les concentrations augmentent tant que l'absorption est plus importante que l'élimination
- Pic des concentrations: à un niveau de concentration donné, la vitesse d'élimination est égale à la vitesse d'absorption et la concentration atteint une valeur maximale,  $C_{max}$ .

- Phase de décroissance des concentrations: les processus d'absorption et d'élimination coexistent toujours mais la vitesse d'élimination est supérieure à la vitesse d'absorption.

L'évolution des concentrations au cours du temps satisfait l'équation :

$$C(t) = \frac{k_a F G_0}{V(k_a - k_e)} (e^{-k_e t} - e^{-k_a t}), \quad (1.8)$$

où

- $F$  est la fraction de la dose administrée au temps zéro,  $G_0$ , qui atteint la circulation sanguine suite à l'administration du médicament par la voie extravasculaire.
- $k_a$  est la constante d'absorption du médicament, dont la valeur dépend du type du médicament, de la forme pharmaceutique (tablette, capsule, solution, poudre, etc.) et de la voie d'administration (orale, intramusculaire, cutanée etc.).

## 1.2 Mesures de biodisponibilité

Dans ce qui suit, nous décrivons quelques paramètres les plus utilisés en pratique. Ces paramètres sont :

- l'exposition de l'organisme au médicament,  $AUC$ ;
- la concentration plasmatique (sanguine) maximale atteinte pendant la phase d'absorption du médicament (traitement),  $C_{max}$ ;
- le temps où la concentration maximale est atteinte,  $t_{max}$ .

Le paramètre pharmacocinétique utilisé pour quantifier l'intensité d'absorption du médicament, est la surface totale sous la courbe des concentrations au cours du temps,  $\int_0^{\infty} C(t) dt$ , couramment notée par  $AUC_{0-\infty}$ . Ce paramètre représente l'exposition totale de l'organisme au médicament, car la surface sous la courbe des concentrations est proportionnelle à la quantité totale de médicament ayant atteint la circulation sanguine.

Pour le modèle mono-compartmental la valeur de  $AUC_{0-\infty}$  s'obtient à partir de la formule :

$$AUC_{0-\infty} = \frac{C_0}{k_e},$$

où  $C_0$  représente la concentration initiale obtenue par une administration intraveineuse du médicament et  $k_e$  est définie par  $k_e = 0.693/t_{1/2}$ .

Par l'approche non-compartmentale, la surface sous la courbe des concentrations est calculée par analyse graphique. En appliquant la méthode d'interpolation qui utilise la règle trapézoïdale (voir appendice A), on peut évaluer le paramètre  $AUC_{0-t}$ , qui représente la surface sous la courbe des concentrations plasmatiques (sanguines) sur l'intervalle de temps  $(0, t)$ . La valeur de  $t$  correspond au temps du dernier prélèvement sanguin.

Soit  $C_0, C_1, \dots, C_k$  les concentrations plasmatiques (sanguines) obtenues aux temps  $0, t_1, \dots, t_k$  respectivement. Alors  $AUC_{0-t_k}$  s'évalue numériquement par la formule suivante :

$$AUC_{0-t_k} = \sum_{i=1}^k \frac{C_{i-1} + C_i}{2} (t_i - t_{i-1}).$$

Notons que la surface sous la courbe des concentrations à partir de  $t_k$  et jusqu'à l'infini pourrait être importante, si la concentration plasmatique (sanguine) au temps  $t_k$  est élevée. La surface de cette dernière portion peut s'évaluer en divisant la concentration du dernier prélèvement,  $C_k$ , par la constante d'élimination du médicament de l'organisme,  $k_e$ . Donc la valeur du  $AUC_{0-\infty}$  s'obtient par:

$$AUC_{0-\infty} = AUC_{0-t_k} + \frac{C_k}{k_e}.$$

En pratique, on observe  $C_0, C_1, \dots, C_k$  pour un patient donné. Alors, un petit nombre de valeurs manquantes et/ou observations aberrantes sur la courbe concentration-temps

(intervalle  $(t_1, t_k)$ ) n'a pas de gros impact sur le calcul du  $AUC_{0-t_k}$  et du  $AUC_{0-\infty}$ . Par contre, s'il existe beaucoup de valeurs manquantes et/ou valeurs inattendues sur la courbe et surtout aux extrémités  $(t_1, t_k)$ , le biais d'estimation du  $AUC$  est substantiel.

La concentration plasmatique (sanguine) maximale atteinte pendant la phase d'absorption du médicament,  $C_{\max}$ , et le temps où cette concentration est atteinte,  $t_{\max}$ , représentent deux paramètres pharmacocinétiques dont les valeurs sont analysées pour caractériser le taux d'absorption du médicament dans l'organisme. Par l'approche non-compartimentale, les valeurs de ces paramètres sont estimées directement à partir de la courbe des concentrations observées. La valeur maximale des concentrations,  $C_{\max} = \max\{C_0, C_1, \dots, C_k\}$ , et  $t_{\max}$  est le temps où cette concentration de médicament est observée.

Pour un modèle mono-compartimental, les valeurs de  $C_{\max}$  et  $t_{\max}$  sont évaluées à partir de la formule (1.8). En développant cette formule on obtient :

$$C(t) = \frac{k_a FG_0}{V(k_a - k_e)} e^{-k_e t} - \frac{k_a FG_0}{V(k_a - k_e)} e^{-k_a t} \quad (1.9)$$

La dérivée  $dC/dt$  de  $C(t)$  donnée par l'équation (1.9) est :

$$\frac{dC}{dt} = \frac{k_a^2 FG_0}{V(k_a - k_e)} e^{-k_a t} - \frac{k_a k_e FG_0}{V(k_a - k_e)} e^{-k_e t}. \quad (1.10)$$

Au moment  $t_{\max}$  où la concentration plasmatique (sanguine) atteint la valeur maximale, la dérivée est  $dC/dt = 0$ . Alors,

$$\frac{k_a^2 FG_0}{V(k_a - k_e)} e^{-k_a t_{\max}} = \frac{k_a k_e FG_0}{V(k_a - k_e)} e^{-k_e t_{\max}} \quad (1.11)$$

qui se réduit à :

$$\frac{k_a}{k_e} = \frac{e^{-k_e t_{\max}}}{e^{-k_a t_{\max}}}. \quad (1.12)$$

En transformant en logarithme naturel les deux membres de (1.12) et en solutionnant pour la valeur  $t_{\max}$ , on obtient :

$$t_{\max} = \frac{\ln(k_a) - \ln(k_e)}{k_a - k_e}, \quad (1.13)$$

ou, en base 10,

$$t_{\max} = \frac{2.303}{k_a - k_e} \log\left(\frac{k_a}{k_e}\right). \quad (1.14)$$

La valeur de  $C_{\max}$  se calcule en substituant la valeur de  $t_{\max}$  dans l'équation (1.8) :

$$C_{\max} = \frac{k_a F G_0}{V(k_a - k_e)} \left( e^{-k_e t_{\max}} - e^{-k_a t_{\max}} \right) \quad (1.15)$$

Pour résumer, les paramètres pharmacocinétiques le plus souvent utilisés en bioéquivalence sont :  $AUC$ ,  $C_{\max}$  et  $t_{\max}$ , qui sont calculés ou estimés soit par l'approche compartimentale soit par l'approche non-compartimentale.

### 1.3 Exemple de calcul et estimation des paramètres pharmacocinétiques

Dans le tableau 1.1 sont donné les concentrations de primidone ( $mg/mL$ ) obtenues suite à l'administration du traitement à un sujet (Chow, 2000). Des prélèvements de sang ont été faits tout de suite après l'administration du traitement et aux heures : 0.5, 1.0, 1.5, 2, 3, 4, 6, 8, 12, 16, 24 et 32.

**Tableau 1.1** Concentrations sanguines de primidone ( $mg/mL$ ) obtenues suite à l'administration du traitement à un sujet.

$i$	$t_i$	$C_i$	$(C_i + C_{i-1})/2$	$t_i - t_{i-1}$	$(C_i + C_{i-1})(t_i - t_{i-1})/2$
1	0.0	0.0			
2	0.5	0.0	0.00	0.5	0.00
3	1.0	2.8	1.40	0.5	0.70
4	1.5	4.4	3.60	0.5	1.80
5	2.0	4.4	4.40	0.5	2.20
6	3.0	4.7	4.55	1.0	4.55
7	4.0	4.1	4.40	1.0	4.40
8	6.0	4.0	4.05	2.0	8.10
9	8.0	3.6	3.80	2.0	7.60
10	12.0	3.0	3.30	4.0	13.20
11	16.0	2.5	2.75	4.0	11.00
12	24.0	2.0	2.25	8.0	18.00
13	32.0	1.6	1.80	8.0	14.40

En ce qui suit, on estime les valeurs de certains paramètres pharmacocinétiques d'après les méthodes du modèle non-compartmental.

Le paramètre pharmacocinétique,  $AUC_{0-32}$ , qui représente l'exposition de l'organisme au médicament sur l'intervalle de temps de 0.0 à 32 heures se calcule de la façon suivante :

$$\begin{aligned}
 AUC_{0-32} &= \sum_{i=1}^{12} \frac{C_{i-1} + C_i}{2} (t_i - t_{i-1}) \\
 &= \frac{0 + 0}{2} (1 - 0.5) + \frac{2.8 + 0}{2} (1.5 - 1) + \dots + \frac{1.6 + 2}{2} (32 - 24) \\
 &= 85.95 (\mu \cdot h/mL).
 \end{aligned}$$

La concentration maximale,  $C_{max} = \max(0, 0, 2.8, \dots, 1.6) = 4.7 \mu g/m$ , a été atteinte à 3.0 heures, c'est-à-dire  $t_{max} = 3.0h$ .

Pour estimer la valeur de la constante d'élimination,  $k_e$ , on considère les dernières 7 valeurs de la concentration du primidone ( $mg/mL$ ) dans le sang du sujet, c'est-à-dire on prend que les concentrations enregistrées après que  $C_{max}$  a été atteint. On transforme ces valeurs en  $\log_{10}$  et on applique la régression linéaire pour déterminer l'équation de cette partie de la courbe, c'est-à-dire de la partie de la courbe qui correspond à l'élimination du traitement du sang. On obtient, donc, l'équation :

$$\log_{10}(C_i) = 0.67134 - 0.01518t_i.$$

Tel que mentionné, à partir de la pente de cette équation,  $\lambda = -0.01518$ , on peut calculer la constante d'élimination du traitement du sang en appliquant la formule :

$$k_e = -2.303\lambda = 2.303 * 0.01518 = 0.03496 \text{ (unités : } h^{-1}\text{)}.$$

À partir de cette valeur on calcule facilement la demi-vie du traitement qui est égale à :

$$t_{1/2} = 0.693/k_e = 0.693/0.03496 = 19.8 \text{ (unités : } h\text{)}.$$

Finalement, le paramètre pharmacocinétique  $AUC_{0-\infty}$  qui représente l'exposition totale de l'organisme au médicament se calcule comme suit :

$$AUC_{0-\infty} = AUC_{0-32} + C_{32}/k_e = 85.95 + 1.6/0.03496 = 131.72 \text{ (unités : } \mu g \cdot h/mL\text{)}.$$

Il faut noter que pour cet exemple la concentration maximale est  $C_{max} = 4.7 \mu g/m$  et que la courbe d'élimination du traitement de l'organisme n'est composée que des dernières 7 concentrations (voir tableau 1.1). Donc, pour estimer la constante d'élimination, on a considéré ces 7 dernières concentrations. Toutefois, en pratique, le nombre de concentrations utilisées pour cette estimation dépend de la courbe des concentrations pour chaque sujet.

Finalement, si on a un échantillon de  $n$  sujets, on répète les calculs du tableau 1.1 pour chacun des  $n$  sujets. Dans ce qui suit, notre référence statistique va porter sur un échantillon de valeurs  $AUC_{0-\infty}$ , noté :  $AUC_1, AUC_2, \dots, AUC_n$ .

## CHAPITRE II

### TESTS EN BIOÉQUIVALENCE

Dans le cadre d'une étude d'efficacité, un nouveau traitement est comparé à un placebo. Le but de cette expérience serait de démontrer que le nouveau traitement donne de meilleurs résultats thérapeutiques que le placebo. En termes de test d'hypothèses (approche traditionnelle) cela peut être décrit par

$$H_0: \mu = 0 \text{ contre } H_a: \mu > 0,$$

où  $\mu = \mu_T - \mu_P$  est la différence entre la moyenne du traitement, appelé Test,  $\mu_T$ , et la moyenne du placebo,  $\mu_P$ . L'hypothèse  $H_0$  représente l'hypothèse par défaut, le statu quo. En d'autres termes, le but de cette expérience serait de démontrer qu'il existe une différence significative entre l'effet du nouveau traitement et l'effet du placebo.

Dans le contexte de la bioéquivalence, le but d'une étude serait de démontrer qu'un nouveau traitement, appelé Test, est bioéquivalent à un traitement standard, appelé Référence. Si l'approche traditionnelle est utilisée pour démontrer la bioéquivalence, l'hypothèse d'équivalence des traitements,  $H_0$ , représenterait l'hypothèse par défaut. Si jamais la vraisemblance de l'hypothèse d'inéquivalence  $H_a$  ne peut pas être prouvée, cela impliquerait que l'expérience n'avait pas été bien planifiée et qu'il n'y a pas eu assez d'évidence pour rejeter l'hypothèse nulle. Un peu plus loin dans ce chapitre on parle plus en détails de la motivation pratique des tests utilisés en bioéquivalence.

On commence, donc, ce chapitre avec un rappel de quelques notions de base en inférence statistique et on décrit brièvement certains tests classiques pour des échantillons

indépendants et appariés. On parle plus en détails du test de permutations et on continue avec la définition et la description des tests d'équivalence.

## 2.1 Généralités sur les tests d'hypothèses

Nous commençons par présenter brièvement des notions de base et des tests classiques de la pratique statistique.

*Définition 2.1 :* Le **test d'hypothèses** représente un processus de dichotomisation des conclusions possibles d'une expérience et l'utilisation de la théorie probabiliste dans le but de choisir une option plutôt qu'une autre.

Les hypothèses concurrentes sont nommées respectivement **l'hypothèse nulle**,  $H_0$ , et **l'hypothèse alternative**,  $H_a$ . Généralement, le choix de l'hypothèse nulle se fait comme dans le cas d'un accusé, qui est considéré innocent tant que sa culpabilité n'a pas été encore prouvée.

*Définition 2.1.a :* **L'erreur de type I** (notée  $\alpha$ ) représente la probabilité de rejeter l'hypothèse nulle dans la faveur de l'hypothèse alternative, dans le cas où l'hypothèse nulle est vraie.

*Définition 2.1.b :* **L'erreur de type II** (notée souvent par  $\beta$ ) représente la probabilité de ne pas rejeter l'hypothèse nulle dans la faveur de l'hypothèse alternative, dans le cas où l'hypothèse nulle est fausse.

*Définition 2.2 :* Un **test statistique** est une fonction des données. La décision de rejeter ou de ne pas rejeter l'hypothèse nulle se base sur les valeurs de cette fonction. La **région critique** représente l'ensemble des valeurs du test statistique pour lesquelles l'hypothèse nulle est rejetée. La **valeur critique** représente le point particulier qui sépare la région de rejet de celle d'acceptation pour une seule variable.

*Définition 2.3 :* La probabilité que la statistique du test se trouve dans la région critique, quand l'hypothèse nulle est vraie, s'appelle **seuil de signification** et il est noté par  $\alpha$ .

*Définition 2.4 :* La **valeur-p** (*p-value*) associée à un test statistique représente la probabilité d'obtenir une valeur du test statistique qui est si extrême ou plus extrême que ce qui a été observée (relativement à l'hypothèse alternative), étant donné que l'hypothèse nulle est vraie.

*Définition 2.5 :* Deux variables aléatoires  $S$  et  $T$  sont appelées **égales en loi**, si elles ont la même fonction de répartition, c'est-à-dire  $P(S \leq x) = P(T \leq x)$  pour tous les  $x$ . La notation utilisée dans ce cas est  $S \stackrel{d}{=} T$ , qui est interprétée comme :  $S$  a la même loi que  $T$ .

## 2.2 Tests classiques de comparaison des moyennes

### 2.2.1 Échantillons indépendants

Dans un plan d'expérience classique, un échantillon de  $n$  sujets reçoit le traitement  $A$  et un échantillon de  $m$  sujets reçoit le traitement  $B$  dans le cadre d'une étude clinique. L'ensemble des données est composé de  $(n + m)$  variables aléatoires indépendantes est  $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$ , où  $X_i$  représente le résultat observé suite à l'administration du traitement  $A$  et  $Y_j$  est le résultat observé suite à l'administration du traitement  $B$ . Dans ce contexte, la différence possible entre les traitements peut être représentée par la relation :

$$X_i \stackrel{d}{=} Y_j + \theta,$$

c'est-à-dire les deux variables aléatoires sont identiques en loi à part un paramètre de position.

Notons que le plan d'expérience utilisé pour cette étude est un plan parallèle. D'après ce plan, chaque sujet est attribué aléatoirement à l'un des traitements. Donc, durant l'étude chaque sujet ne reçoit qu'un seul traitement. Plus de détails sur ce plan d'expérience sont donnés au chapitre 3.

Dans le cadre du plan d'expérience parallèle, les traitements sont égales en loi, c'est-à-dire la loi du traitement  $A$  peut être obtenue en décalant toutes les valeurs du traitement  $B$

par la même distance  $|\theta|$  à droite (si  $\theta > 0$ ) ou à gauche (si  $\theta < 0$ ). Certaines conditions doivent être satisfaites :

1. Les variables aléatoires  $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$ , sont indépendantes. Les  $X_i$ ,  $i = 1, \dots, n$  sont de même fonction de répartition,  $F$  et les  $Y_j$ ,  $j = 1, 2, \dots, m$  sont de même fonction de répartition,  $G$ .
2. Il existe une constante  $\theta \in R$ , telle que  $F(x) = G(x - \theta)$  pour tout  $x \in R$ .

En effet, si on suppose que  $X_i \stackrel{d}{=} Y_j + \theta$ , les fonctions de répartition  $F_{X_i}(x) = G_{Y_j + \theta}(x)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ . D'après la définition d'une fonction de répartition on obtient  $P(X_i \leq x) = P(Y_j + \theta \leq x) = P(Y_j \leq x - \theta)$ . Cela implique l'égalité  $F_{X_i}(x) = G_{Y_j}(x - \theta)$  (condition 2).

Les tests classiques dans le contexte des échantillons indépendants sont :

1. le test t de Student;
2. le test de Wilcoxon pour deux échantillons indépendants;
3. le test non paramétrique de permutations qui prend de plus en plus de place en ce qui concerne son application.

En ce qui suit on parle brièvement du test t de Student pour deux échantillons indépendants et du test de Wilcoxon pour deux échantillons indépendants. On donne plus de détails sur le test de permutations.

### 1. Test t de Student pour deux échantillons indépendants

Soit  $X_1, X_2, \dots, X_n$  et  $Y_1, Y_2, \dots, Y_m$  deux échantillons aléatoires indépendants. De plus, on suppose que ces variables suivent des lois normales des moyennes  $\mu_A$  et  $\mu_B$ , respectivement, et de même variance  $\sigma^2$ .

On veut tester l'hypothèse :

$$H_0 : \mu_A - \mu_B = \theta_0$$

contre l'une des alternatives

$$H_{a1} : \mu_A - \mu_B \neq \theta_0,$$

$$H_{a2} : \mu_A - \mu_B > \theta_0,$$

$$H_{a3} : \mu_A - \mu_B < \theta_0.$$

Le plus souvent  $\theta_0 = 0$ .

Lorsque la variance est inconnue et les tailles des échantillons sont grandes, la statistique du test au seuil de signification  $\alpha\%$  est définie par :

$$Z_{\bar{X}-\bar{Y}} = \frac{\bar{X}-\bar{Y}-\theta}{\sqrt{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)}}$$

où  $\bar{X} - \bar{Y}$  suis une loi normale de moyenne  $\theta$  et de variance  $\sigma^2(1/n + 1/m)$  sous l'hypothèse nulle et  $s_X^2 = 1/(n-1) \sum_{i=1}^n (X_i - \bar{X})^2$  et  $s_Y^2 = 1/(m-1) \sum_{j=1}^m (Y_j - \bar{Y})^2$ .

Par contre, si les tailles des échantillons sont petites, la statistique du test au seuil de signification de  $\alpha\%$  suit une loi t de Student avec  $n + m - 2$  degrés de liberté et elle est définie par :

$$T_{\bar{X}-\bar{Y}} = \frac{\bar{X}-\bar{Y}-\theta}{S_p \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

$$\text{où } S_p = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}.$$

## 2. Test de Wilcoxon pour deux échantillons

Un traitement  $A$  est administré aléatoirement à  $n$  sujets et un traitement  $B$  est administré aléatoirement à  $m$  sujets. Les résultats obtenus sont rangés en ordre croissant en leur assignant un rang  $R_1, R_2, \dots, R_N$ ,  $N = n + m$ . Si deux ou plusieurs valeurs sont égales («ties»), on leur donne un **rang moyen**, défini par la moyenne des rangs que ces valeurs auraient eu si elles étaient différentes.

On prend, comme précédemment,  $\theta = \mu_A - \mu_B$  et on propose de tester l'hypothèse

$$H_0 : \theta = 0$$

contre l'une des hypothèses alternatives :

$$H_a : \theta \neq 0, \text{ test bilatéral}$$

$$H_a : \theta > 0 \text{ ou } H_a : \theta < 0, \text{ tests unilatéraux.}$$

Considérons par exemple le cas, où l'hypothèse alternative est :

$$H_a : \theta > 0.$$

L'hypothèse  $H_0$  est rejetée si les résultats obtenus suite à l'administration du traitement  $A$  reçoivent des rangs assez élevés. La statistique du test est définie par :

$$V_s = R_1 + R_2 + \dots + R_n$$

L'hypothèse  $H_0$  est rejetée au seuil de signification de  $\alpha$  et, donc, le traitement  $A$  est considéré plus efficace par rapport au traitement  $B$ , si

$$V_s \geq c,$$

où  $c$  est la valeur critique qui satisfait l'égalité suivante:

$$P_{H_0}(V_s \geq c) = \alpha.$$

La valeur critique est établie à partir du résultat que, sous  $H_0$ , la loi de toutes les répartitions des rangs  $R_1, R_2, \dots, R_N$  à travers ces  $N = n + m$  sujets est uniforme sur  $\{1, 2, \dots, N\}$ . En d'autres mots, si  $\theta = 0$  il n'y a pas de raison de supposer que les personnes qui ont reçu le traitement ont des réponses de rangs plus élevés parmi les  $N$ .

### 3. Test des permutations

Généralement, les études de bioéquivalence sont basées sur des échantillons de petite taille dont les sujets sont attribués aléatoirement aux traitements. En pratique, on utilise le test *t* de Student pour réaliser l'inférence statistique dans le cadre de ces études. Généralement, l'une des hypothèses importantes sous lesquelles ce test paramétrique est appliqué est la normalité de la loi des observations ou d'une transformation de celle-ci. De plus, les observations doivent être indépendantes et leurs variances égales. Si jamais, l'une de ces hypothèses n'est pas satisfaite, les probabilités obtenues en appliquant le test paramétrique ne sont plus exactes. Cependant, le test de permutations n'exige aucune hypothèse sur la loi des données, à part le fait que les sujets soient attribués aléatoirement aux traitements. En recalculant la statistique pour toutes les permutations possibles de l'échantillon observé, le test de permutations génère lui-même la loi de référence.

Soit un exemple où un nouveau traitement  $A$  est comparé à un traitement standard  $B$  (Ernst, 2004). Les observations constituent le temps (en jours) de rétablissement post opératoire des sujets.  $N$  sujets ont participé à cette étude, parmi lesquels  $n$  ont reçu le nouveau traitement  $A$  et  $m = N - n$  ont reçu le traitement standard  $B$ .

Considérons un exemple où les hypothèses d'intérêt sont :

$$H_0: \mu_A - \mu_B = 0 \text{ contre } H_a: \mu_A - \mu_B < 0.$$

Notons que d'autres hypothèses alternatives peuvent aussi être définies :

$$H_{a1}: \mu_A - \mu_B > 0,$$

$$H_{a2}: \mu_A - \mu_B \neq 0.$$

Soit noté par  $X_1, X_2, \dots, X_n$  et  $Y_1, Y_2, \dots, Y_m$  les résultats obtenus suite à l'administration du traitement  $A$  et  $B$ , respectivement. Pour mesurer la différence entre les traitements, la valeur  $D = \bar{X} - \bar{Y}$  est calculée. On veut déterminer si cette différence est assez extrême par rapport à une certaine loi de référence pour que le nouveau traitement soit considéré meilleur.

Dans le tableau 2.1 sont donnés les résultats obtenus suite à l'administration du nouveau traitement  $A$  à  $n = 4$  sujets et du traitement standard  $B$  à  $m = 3$  sujets. La différence entre les moyennes observées est de  $\bar{X} - \bar{Y} = -9$  jours, qui indique une possible diminution du temps de rétablissement après l'administration du traitement  $A$ .

Si jamais l'hypothèse nulle est vraie, c'est-à-dire qu'il n'existe pas de différence significative entre les traitements, alors le temps de rétablissement du chaque sujet est le même indépendamment du traitement reçu. Par exemple, le sujet 2 s'est rétabli en 22 jours après avoir reçu le nouveau traitement  $A$ . Sous l'hypothèse nulle, ça lui aurait pris la même période de temps pour se rétablir, si on lui administrait le traitement standard  $B$ . Donc, le temps de rétablissement n'est pas aléatoire, car les sujets ne sont pas tirés aléatoirement d'une population dans cet exemple. Par contre, l'attribution des sujets aux traitements est aléatoire. La loi discrète de probabilité de  $\bar{X} - \bar{Y}$  peut être construite en se basant sur l'attribution aléatoire des  $N$  sujets aux deux traitements. L'ensemble des valeurs observées ne constitue qu'une seule permutation parmi la totalité de  $\binom{N}{n}$  permutations équiprobables, où la probabilité d'observer une valeur parmi  $\binom{N}{n}$  est de  $1/\binom{N}{n}$ . Donc la loi de  $\bar{X} - \bar{Y}$  peut être construite en calculant la valeur  $\bar{X} - \bar{Y}$  pour toutes les permutations possibles de sujets au traitement  $A$  et au traitement  $B$ . Dans le tableau 2.1 sont énumérées toutes les permutations possibles pour la totalité de 7 résultats, c'est-à-dire des 4 observations obtenues pour le traitement  $A$  et 3 observations obtenues pour le traitement  $B$ .

**Tableau 2.1** Toutes les permutations possibles de 7 résultats obtenus suite à l'administration du traitement *A* à  $n = 4$  sujets et du traitement *B* à  $m = 3$  sujets.

Permutation	Traitement							Différence des moyennes $\bar{X} - \bar{Y}$
	<i>A</i>				<i>B</i>			
1	19	22	25	26	23	33	40	-9
2	22	23	25	26	19	33	40	-6.67
3	22	33	25	26	19	23	40	-0.83
4	22	25	26	40	19	23	33	3.25
5	19	23	25	26	22	33	40	-8.42
6	19	25	26	33	22	23	40	-2.58
7	19	25	26	40	22	23	33	1.50
8	19	22	23	26	25	33	40	-10.17
9	19	22	26	33	23	25	40	-4.33
10	19	22	26	40	23	25	33	-0.25
11	19	22	23	25	26	33	40	-10.75
12	19	22	25	33	23	26	40	-4.92
13	19	22	25	40	23	26	33	-0.83
14	23	25	26	33	19	22	40	-0.25
15	22	23	26	33	19	25	40	-2.00
16	22	23	25	33	19	26	40	-2.58
17	19	23	26	33	22	25	40	-3.75
18	19	23	25	33	22	26	40	-4.33
19	19	22	23	33	25	26	40	-6.08
20	23	25	26	40	19	22	33	3.83
21	22	23	26	40	19	25	33	2.08
22	22	23	25	40	19	26	33	1.50
23	19	23	26	40	22	25	33	0.33
24	19	23	25	40	22	26	33	-0.25
25	19	22	23	40	25	26	33	-2.00
26	25	26	33	40	19	22	23	9.67
27	22	26	33	40	19	23	25	7.92
28	22	25	33	40	19	23	26	7.33
29	19	26	33	40	22	23	25	6.17
30	19	25	33	40	22	23	26	5.58
31	19	22	33	40	23	25	26	3.83
32	23	26	33	40	19	22	25	8.50
33	23	25	33	40	19	22	26	7.92
34	22	23	33	40	19	25	26	6.17
35	19	23	33	40	22	25	26	4.42

Dans ce tableau sont aussi présentées les différences entre les moyennes calculées pour chacune des permutations. Sous l'hypothèse nulle, la probabilité d'une permutation est égale à  $1/35$  pour cet exemple.

En conclusion, le test de permutations permet de déterminer la loi de référence en calculant la statistique du test pour toutes les attributions possibles des sujets aux traitements  $A$  et  $B$ .

Notons que la statistique considérée dans cet exemple représente la différence entre les moyennes des observations obtenues pour les deux traitements. C'est la même statistique que celle utilisée dans le cadre des études de bioéquivalence. Cependant, d'autres statistiques peuvent être considérées pour le test de permutation, comme par exemple :

- la différence des médianes ;
- la somme des résultats obtenus suite à l'administration d'un des traitements ;
- la somme des rangs des observations d'un des traitements après que les résultats des deux traitements soient rangés et qu'ils aient reçu un rang de 1 à  $N$ .

La dernière statistique est celle du test de Wilcoxon pour deux échantillons. Donc, ce test est un cas particulier du test de permutations.

La valeur-p du test de permutations est la probabilité d'obtenir une statistique extrême ou plus extrême que la valeur  $d^* = \bar{X}^* - \bar{Y}^*$  observée de l'échantillon. Étant donné que toutes les permutations  $\binom{N}{n}$  sont équiprobables, sous l'hypothèse nulle, la valeur-p est définie par :

$$p = P(D \leq d^* | H_0) = \frac{\sum_{i=1}^{\binom{N}{n}} I(d_i \leq d^*)}{\binom{N}{n}},$$

où  $d_i$  c'est la valeur de la statistique  $D = \bar{X} - \bar{Y}$  pour la permutation  $i$  et  $I(\cdot)$  est la fonction indicatrice. Pour l'exemple présenté dans le tableau 2.1, la statistique observée est  $d^* = -9$  et la valeur-p est la suivante :

$$p = P(\bar{X} - \bar{Y} \leq -9) = \frac{3}{35} \approx 0.0857.$$

D'après ce résultat on constate qu'il n'existe pas de différence significative entre les traitements  $A$  et  $B$  au seuil de 5%.

Notons que dans ce cas les valeurs-p doivent être des multiples de la probabilité  $1/\binom{N}{n}$ , c'est-à-dire elles sont égales à l'une des valeurs  $k/\binom{N}{n}$ , où  $k = 1, 2, \dots, \binom{N}{n}$ . Si le seuil de signification  $\alpha = k_0/\binom{N}{n}$  pour un  $k_0$ , alors

$$P(\text{Erreur type I}) = P\left(\sum_{i=1}^{\binom{N}{n}} I(d_i \leq d^*) \leq k_0 \mid H_0\right) = \frac{k_0}{\binom{N}{n}} = \alpha. \quad (2.1)$$

Donc, le test de permutations est un test exact. Si jamais, la valeur de  $\alpha$  est considérée égale à l'un des  $k/\binom{N}{n}$ ,  $k = 1, 2, \dots, \binom{N}{n}$ , alors on retient la valeur  $k$  telle que

$$P(\text{Erreur type I}) = \frac{k}{\binom{N}{n}} < \alpha$$

et le test de permutations est conservateur. Toutefois, dans les deux cas, le test de permutations assure le contrôle de l'erreur de type I avec une seule condition : que les sujets soient attribués aléatoirement aux traitements.

Le test de permutations peut être réalisé d'après les étapes suivantes (pratique courante) :

- on calcule la statistique de l'ensemble observé,  $d^*$  ;
- on permute l'ensemble observé de  $\binom{N}{n}$  façons équiprobables et on calcule la statistique du test  $d_i$  pour chacun des ensembles permutés ;
- on calcule la valeur-p d'après la formule (2.1).

Ce test est utilisé dans notre étude de simulation (chapitre 4).

### 2.2.2 Observations appariées

Dans un autre plan d'expérience classique, chaque sujet reçoit deux traitements  $A$  et  $B$ . Soit  $(X, Y)$  la paire aléatoire, où  $X$  et  $Y$  représentent les résultats obtenus suite à l'administration des traitements  $A$  et  $B$ , respectivement. Ces données sont appelées

appariées. Les conditions sous lesquelles les deux traitements ont été administrés sont considérées strictement équilibrées. Cela permet d'interpréter la différence intra-sujet  $D = X - Y$  qui quantifie l'efficacité du traitement  $A$  par rapport au traitement  $B$ , pour le même sujet. Dans le contexte des données appariées, la notion de **positionnement** est introduite par une nouvelle hypothèse par laquelle toute différence potentielle entre les traitements  $A$  et  $B$  représente le déplacement de la loi de la différence  $D$  de  $\theta$  unités par rapport à l'origine zéro. Par contre, la loi dans ce cas est la même pour les deux traitements.

Dans l'absence d'une différence entre les traitements, on suppose que la loi de la différence  $D$  est de fonction de densité  $f_0 : R \rightarrow [0,1]$ , continue et symétrique par rapport au zéro.

L'ensemble  $(X_1, Y_1), \dots, (X_N, Y_N)$  représente un échantillon aléatoire de toutes les  $N$  paires d'observations obtenues suite à l'administration des deux traitements. L'inférence statistique est alors basée sur des hypothèses qui sont définies par rapport aux différences  $D_i = X_i - Y_i$ . Dans ce cas les conditions suivantes doivent être satisfaites :

1.  $D_1, D_2, \dots, D_N$  sont indépendantes et identiquement distribuées (i.i.d.).
2.  $P[D_i \leq d] = F_0(d - \theta)$ , pour tout  $i = 1, 2, \dots, N$ , où  $F_0(-d) = 1 - F_0(d)$  et  $d, \theta$  représente le paramètre d'intérêt.

Notons que les valeurs positives  $\theta > 0$  indiquent une meilleure efficacité du traitement  $A$  par rapport au traitement  $B$ .

Les tests les plus utilisés dans le contexte des données appariées sont :

1. le test t de Student;
2. le test des rangs signés de Wilcoxon.

En ce qui suit on parle brièvement du test t de Student pour deux échantillons appariés et du test des rangs signés de Wilcoxon.

### 1. Le test t de Student pour données appariées

La comparaison de la moyenne des deux échantillons appariés est fondée sur l'analyse des différences observées pour chacune des  $N$  paires d'observations  $i$ , c'est-à-dire sur  $D_i = X_i - Y_i, i = 1, 2, \dots, N$ .

Lorsque les variables aléatoires  $X_i$  et  $Y_i$  suivent des lois normales de variance inconnue, la statistique du test sous l'hypothèse nulle suit une loi t de Student à  $n - 1$  degrés de liberté et elle se définit par :

$$T_D = \frac{\bar{D}}{s_D/\sqrt{N}}$$

### 2. Le test des rangs signés de Wilcoxon

Le traitement  $A$  et le traitement  $B$  est administré de façon aléatoire à chacune de  $N$  paires de sujets. Notons par  $D_i$  la différence entre le résultat obtenu suite à l'administration du traitement  $A$ ,  $X_i$ , et le résultat obtenu suite à l'administration du traitement  $B$ ,  $Y_i$ ,  $i = 1, 2, \dots, N$ . Considérons le nombre de différences positives égal à  $n$  et le nombre de différences négatives égal à  $m = N - n$ . Le paramètre d'intérêt est la différence des moyennes  $\theta = \mu_A - \mu_B$ , où  $\theta$  est un paramètre de positionnement.

On veut tester l'hypothèse nulle

$$H_0 : \theta = 0$$

contre l'une des hypothèses alternatives :

$$H_a : \theta \neq 0, \text{ test bilatéral,}$$

$$H_a : \theta > 0 \text{ ou } H_a : \theta < 0, \text{ tests unilatéraux.}$$

Pour illustrer ce test considérons le cas où  $H_a : \theta > 0$ .

La statistique du test est obtenue selon les étapes suivantes (pratique courante):

- Étape 1.** On ignore les différences nulles, car elles n'apportent pas d'information sur la différence entre le traitement  $A$  et le traitement  $B$ . On diminue respectivement le nombre  $N$  par le nombre de différences nulles.
- Étape 2.** On range les différences  $|D_i|$  en ordre croissant.
- Étape 3.** On assigne les rangs  $R_1^+ < R_2^+ < \dots < R_n^+$  aux différences positives et les rangs  $R_1^- < R_2^- < \dots < R_m^-$  aux différences négatives. Si deux ou plusieurs valeurs sont égales (*ties*), on leur donne un **rang moyen**, défini par la moyenne des rangs que ces valeurs auraient si elles étaient différentes.
- Étape 4.** On calcule la statistique du test  $V_s^+ = R_1^+ + R_2^+ + \dots + R_n^+$  qui est basée sur la somme des rangs attribués aux différences positives.

On constate que le test des rangs signés tient compte non seulement des signes, mais aussi la valeur des différences  $D_i$ .

L'hypothèse nulle est rejetée lorsque la valeur  $V_s^+$  est suffisamment grande, c'est-à-dire si  $V_s^+ \geq c$ . Il existe des tables de Wilcoxon qui donnent la probabilité d'une valeur au moins aussi grande que la valeur critique  $c$ , au seuil de signification de  $\alpha$ , c'est-à-dire  $P_H(V_s^+ \geq c)$ .

On voit que ce test ressemble au test de Wilcoxon pour deux échantillons, où le premier échantillon serait celui des différences positives, le deuxième celui des différences négatives. Sous  $H_0$ , on a la répartition uniforme sur  $\{1, 2, \dots, N\}$  des  $N$  rangs des «+» et des «-» (conditionnellement au nombre de «+» et des «-»). De plus, la répartition entre «+» et «-» est indépendante de la grandeur relative des différences  $|X_i - Y_i|$ ,  $i = 1, 2, \dots, N$ .

Lorsque  $N$  est grand, la loi de Wilcoxon peut être remplacée par l'approximation normale  $N(\mu, \sigma^2)$ , de paramètres :

$$\mu = E_0(V_s^+) = \frac{N(N+1)}{4},$$

$$\sigma^2 = \text{Var}_0(V_s^+) = \frac{N(N+1)(2N+1)}{24},$$

c'est-à-dire on peut utiliser

$$P_{H_0}(V_s^+ \geq c) \approx \Phi\left(\frac{V_s^+ - E_0(V_s^+) + 0.5}{\sqrt{\text{Var}_0(V_s^+)}}\right).$$

### 2.3 Motivation pratique des tests utilisés en bioéquivalence

Selon Wellek (2000) chapitre 1, dans les problèmes des tests d'hypothèses il existe une asymétrie logique implicite en ce qui concerne les rôles joués par l'hypothèse nulle et l'hypothèse alternative. C'est entre ces deux hypothèses qu'il faut faire un choix en se basant sur des données collectées pendant une étude ou une expérience. Le risque de prendre une décision erronée en faveur de l'hypothèse alternative ne dépasse pas un seuil de signification  $\alpha$  défini d'avance, tandis que le risque de faire un mauvais choix dans la faveur de l'hypothèse nulle peut être aussi élevé que  $(1 - \alpha)$ . En d'autres mots, l'erreur de type I, qui est fixée d'avance, est généralement contrôlée. Cela implique que l'acceptation de l'hypothèse alternative, quand elle est en réalité fausse est contrôlée d'une certaine façon. Si jamais  $\alpha$  est fixé à 5%, le risque de rejeter l'hypothèse nulle quand elle est en réalité vraie (donc d'accepter l'hypothèse alternative quand elle est fausse) ne dépasse pas cette limite. Cependant, l'erreur de type II n'est pas contrôlée. Le risque  $\beta$  peut aller jusqu'à  $1 - \alpha$  (95% si  $\alpha$  est de 5%). La question serait : pourquoi ne pas inverser les hypothèses dans le cadre des études de bioéquivalence pour pouvoir contrôler le risque d'accepter l'hypothèse de bioéquivalence des traitements, dans le cas où cette hypothèse n'est pas vraie en réalité. C'est-à-dire, on propose de tester l'hypothèse de bioinéquivalence,  $H_0$ , pour que le risque de rejeter celle-ci en faveur de l'hypothèse de bioéquivalence,  $H_a$ , dans le cas où l'hypothèse nulle est vraie ne soit pas plus élevé que le seuil de signification  $\alpha$  fixé a priori.

Tout cela suggère l'importance d'inversion des hypothèses dans le cadre des études de bioéquivalence. Pour illustrer cette idée, soit  $\mu_T$  la moyenne d'un traitement Test et  $\mu_R$  la moyenne d'un traitement Référence. Dans ce contexte, notons par  $\theta$  la différence  $\mu_T - \mu_R$  et par  $\theta_0$  une valeur fixée. En vertu de ce principe, on inverse l'hypothèse nulle avec l'hypothèse alternative bilatérale et on obtient :

$$H_0 : \theta \neq \theta_0 \text{ contre } H_a : \theta = \theta_0.$$

Cela nous amène au point essentiel est le suivant : ce n'est pas tout à fait approprié de définir l'hypothèse alternative comme une égalité, car cette définition n'admet pas de solutions.

Afin d'expliquer la difficulté, soit un échantillon aléatoire de  $N$  données  $D_1, D_2, \dots, D_N$  indépendantes et identiquement distribuées tel que défini dans la section 2.2.1 par exemple. Pour la simplicité, on suppose que les données suivent une loi normale  $\mathcal{N}(\mu, \sigma_0)$ , où  $\theta = \mu_T - \mu_R$  et  $\sigma_0$  est connu.

Soit  $\theta_0$  fixé tel que  $\theta \neq \theta_0$ , alors, la probabilité de la région d'acceptation serait égale à :

$$P\left(\bar{X} \geq \theta_0 + z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}} \mid \mu > \theta_0\right) + P\left(\bar{X} \leq \theta_0 - z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}} \mid \mu < \theta_0\right)$$

ce qui donne

$$P\left(\frac{\bar{X} - \theta}{\sigma_0/\sqrt{n}} + \frac{\theta - \theta_0}{\sigma_0/\sqrt{n}} \geq z_{\alpha/2} \mid \theta > \theta_0\right) + P\left(\frac{\bar{X} - \theta}{\sigma_0/\sqrt{n}} + \frac{\theta - \theta_0}{\sigma_0/\sqrt{n}} \leq -z_{\alpha/2} \mid \theta < \theta_0\right),$$

d'où

$$P\left(\frac{\bar{X} - \theta}{\sigma_0/\sqrt{n}} + \Delta \geq z_{\alpha/2} \mid \theta > \theta_0\right) + P\left(\frac{\bar{X} - \theta}{\sigma_0/\sqrt{n}} + \Delta \leq -z_{\alpha/2} \mid \theta < \theta_0\right),$$

où  $\Delta = (\theta - \theta_0)/(\sigma_0/\sqrt{n})$ . Donc,

$$P\left(\frac{\bar{X} - \theta}{\sigma_0/\sqrt{n}} \geq z_{\alpha/2} - \Delta \mid \theta > \theta_0\right) + P\left(\frac{\bar{X} - \theta}{\sigma_0/\sqrt{n}} \leq -z_{\alpha/2} - \Delta \mid \theta < \theta_0\right) =$$

$$P(Z \geq z_{\alpha/2} - \Delta) + P(Z \leq -z_{\alpha/2} - \Delta).$$

La variable aléatoire  $Z$  suit une loi normale  $\mathcal{N}(0,1)$  et  $\Delta > 0$ , si  $\theta > \theta_0$  et  $\Delta < 0$ , si  $\theta < \theta_0$ . Donc on a obtenu

$$1 - \Phi(z_{\alpha/2} - \Delta) + \Phi(-z_{\alpha/2} - \Delta),$$

qui est une fonction continue en  $\Delta$ . Par définition de la continuité, la probabilité va toujours converger vers la valeur  $\alpha$  quand  $\Delta \rightarrow 0$ , donc la puissance est toujours plus grande ou égale à la valeur fixe  $1 - \alpha$ , ce qui est absurde.

Une solution à ce problème serait de définir un intervalle (ou une région) de valeurs  $\theta$  assez proches de la valeur  $\theta_0$  pour que la différence entre  $\theta$  et  $\theta_0$  soit non pertinente de point de vue pratique. Pour conclure, dans le cadre d'une étude d'équivalence, l'idée ne serait pas que d'inverser les hypothèses d'un test traditionnel, mais également de les transformer.

#### 2.4 Tests d'équivalence

Rappelons que dans le cadre d'une étude d'équivalence, on compare un traitement Test, abrégé par  $T$ , et un traitement Référence, abrégé par  $R$ . Soit  $\theta = \mu_T - \mu_R$  le paramètre qui mesure la différence entre les effets de ces traitements. Le test d'équivalence est défini par les hypothèses:

$$H_0 : \theta \leq \theta_0 - \varepsilon_1 \text{ ou } \theta \geq \theta_0 + \varepsilon_2 \quad (2.2)$$

$$\text{contre } H_a : \theta_0 - \varepsilon_1 < \theta < \theta_0 + \varepsilon_2,$$

où  $(\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2)$  représente l'intervalle d'équivalence. Les valeurs  $\varepsilon_1, \varepsilon_2$  sont choisies sans que les données à analyser soient connues.  $H_0$  représente l'hypothèse d'inéquivalence et  $H_a$  représente l'hypothèse d'équivalence.

#### 2.4.1 Test unilatéral (non infériorité)

Comparativement aux études d'équivalence définies par les hypothèses 2.2, dans le cadre d'un test unilatéral d'équivalence la région de rejet n'est bornée qu'inférieurement. C'est-à-dire, par un test unilatéral d'équivalence on vérifie si l'effet du traitement Test,  $T$ , n'est pas significativement inférieur à l'effet du traitement Référence,  $R$ . Cependant, la supériorité de l'effet du traitement  $T$  par rapport à l'effet du traitement  $R$  n'est pas testée. Cela est en contradiction avec les études d'équivalence bilatérale, définies par les hypothèses (2.2), où on vérifie l'hypothèse de non infériorité et de non supériorité du traitement Test par rapport au traitement Référence. La borne supérieure pour un test de non infériorité est de  $+\infty$  comparativement à  $\theta_0 + \varepsilon_2$  d'un test d'équivalence bilatéral.

Les hypothèses du test de non infériorité sont définies par :

$$H_0 : \theta \leq \theta_0 - \varepsilon \text{ contre } H_a : \theta > \theta_0 - \varepsilon, \quad (2.3)$$

où  $\varepsilon > 0$  est suffisamment petit et connu.

Soit les hypothèses d'un test unilatéral ordinaire défini par :

$$H_0 : \theta \leq \theta_0 \text{ contre } H_a : \theta > \theta_0. \quad (2.4)$$

Un test de non infériorité se distingue d'un test unilatéral ordinaire par le déplacement de  $-\varepsilon$  unités de la borne qui sépare les deux hypothèses,  $\theta_0$ .

Même si le test unilatéral d'équivalence se distingue du test unilatéral ordinaire par la définition des hypothèses, de point de vue statistique, juste une légère modification du test

statistique est requise et les conclusions sont faites un peu différemment, mais le principe de raisonnement reste le même. En ce qui suit, cette modification sera décrite plus en détail pour la comparaison des deux traitements,  $T$  et  $R$ , dans le cas des observations appariées et indépendantes.

### (i) Échantillons appariés

Soit le test unilatéral ordinaire défini par les hypothèses :

$$H_0 : \theta \leq 0 \text{ contre } H_a : \theta > 0. \quad (2.5)$$

Ce test peut être obtenu en remplaçant  $\theta_0 = 0$  et  $\varepsilon = 0$  dans (2.3).

Dans le cas des données appariées (voir section 2.2.2), la région de rejet au seuil de signification  $\alpha$  est définie par

$$\{T(D_1, \dots, D_n) > c_\alpha\}, \quad (2.6)$$

où  $T(\cdot)$  est un test statistique et  $c_\alpha$  c'est la  $\alpha$  100% percentile supérieur de la loi des variables aléatoires  $(D_1, D_2, \dots, D_n)$  dans le cas où  $\theta = 0$ .

Si on redéfinit les différences intra-sujet  $D_i$  par  $\tilde{D}_i = D_i + \varepsilon$ , les valeurs du vecteur  $(\tilde{D}_1, \tilde{D}_2, \dots, \tilde{D}_n)$  seront également indépendantes et identiquement distribuées et elles vont satisfaire la condition

$$P(\tilde{D}_i \leq d) = F_0(d - \tilde{\theta})$$

pour  $\tilde{\theta} = \theta + \varepsilon$ , où  $\tilde{\theta}$  représente le paramètre transformé. Pour le paramètre transformé  $\tilde{\theta}$  le test unilatéral ordinaire sera défini par

$$H_0 : \tilde{\theta} \leq 0 \text{ contre } H_a : \tilde{\theta} > 0.$$

Étant donné que  $\tilde{\theta} = \theta + \varepsilon$ , ce test peut être écrit comme

$$H_0 : \theta + \varepsilon \leq 0 \text{ contre } H_a : \theta + \varepsilon > 0,$$

d'où le test d'hypothèses

$$H_0 : \theta \leq -\varepsilon \text{ contre } H_a : \theta > -\varepsilon. \quad (2.7)$$

Le test obtenu de cette façon va rejeter l'hypothèse d'infériorité  $H_0 : \theta \leq -\varepsilon$  si et seulement si la statistique

$$T(D_1 + \varepsilon, \dots, D_n + \varepsilon) > c_\alpha, \quad (2.8)$$

où la fonction  $T(\cdot)$  et  $c_\alpha$  sont calculés de la même façon que dans le cas du test unilatéral ordinaire. L'approche décrite plus haut est illustrée par l'exemple 2.2 donné à la fin de cette section.

## (ii) Échantillons indépendants

Dans le cas des échantillons indépendants, le test unilatéral d'équivalence

$$H_0 : \theta \leq -\varepsilon \text{ contre } H_a : \theta > -\varepsilon \quad (2.9)$$

(cas où  $\theta_0 = 0$ ) peut être réduit au test unilatéral ordinaire

$$H_0 : \tilde{\theta} \leq 0 \text{ contre } H_a : \tilde{\theta} > 0 \quad (2.10)$$

de façon similaire au cas des données appariés. Pour le test ordinaire la région de rejet au seuil de signification  $\alpha$  sera définie par

$$\{T(X_1, \dots, X_n, Y_1, \dots, Y_m) > c_\alpha\}, \quad (2.11)$$

où  $T(\cdot)$  est une statistique de  $n + m$  arguments,  $c_\alpha$  c'est la  $\alpha$  100% percentile supérieur de la loi des variables aléatoires  $(X_1, \dots, X_n, Y_1, \dots, Y_m)$  quand  $\theta = 0$ .

Si  $X_i$  est transformé en  $\tilde{X}_i = X_i + \varepsilon$ ,  $i = 1, \dots, n$  et on pose  $\tilde{Y}_j = Y_j$ ,  $j = 1, \dots, m$ , alors  $\tilde{X}_i \stackrel{d}{=} \tilde{Y}_j + \theta$ .

En effet, comme  $X_i \stackrel{d}{=} Y_j + \theta$ , alors  $P(X_i \leq x) = P(Y_j + \theta \leq x)$ . Pour une certaine valeur  $\varepsilon$ , les probabilités suivantes sont égales

$$P(X_i + \varepsilon \leq x + \varepsilon) = P(Y_j + \theta \leq x),$$

d'où on constate que

$$F_{X_i + \varepsilon}(x + \varepsilon) = G_{Y_j + \theta}(x)$$

et donc

$$F_{\tilde{X}_i}(\tilde{x}) = G_{\tilde{Y}_j}(\tilde{x} - \theta).$$

Alors,

$$\tilde{X}_i \stackrel{d}{=} \tilde{Y}_j + \theta,$$

c'est à dire les variables aléatoires  $\tilde{X}_i$  et  $\tilde{Y}_j + \theta$  sont égales en loi.

Pour les variables aléatoires transformées les suppositions suivantes sont satisfaites.

On a :

- les variables  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n, \tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_m$  sont indépendantes, tous les  $\tilde{X}_i$  sont de même fonction de répartition  $F$ , et tous les  $\tilde{Y}_j$  sont de même fonction répartition  $G$
- pour une constante  $\theta \in R$ ,  $F(\tilde{x}) = G(\tilde{x} - \theta)$  pour tout  $\tilde{x} \in R$ .

De façon similaire au cas des données appariées, le test statistique pour les variables transformées peut se définir par

$$T(X_1 + \varepsilon, \dots, X_n + \varepsilon, Y_1, \dots, Y_m) > c_\alpha, \quad (2.12)$$

où la statistique  $T(\cdot)$  et  $c_\alpha$  sont calculés de la même façon que dans le cas du test unilatéral ordinaire.

Il faut noter que l'approche paramétrique peut être facilement implémentée dans ce cas, si on considère les fonctions de répartition

$$F(x) = \Phi\left(\frac{x-\theta}{\sigma}\right) \text{ et } G(y) = \Phi\left(\frac{y}{\sigma}\right),$$

respectivement, qui représentent des lois normales avec la même variance,  $\sigma^2$ . Alors les deux échantillons peuvent être comparés à l'aide du test t de Student. Sous l'hypothèse nulle, la statistique  $T(\cdot)$  se définit par :

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{1/n_x + 1/n_y}}, \quad (2.13)$$

où la différence  $\bar{X} - \bar{Y}$  doit être changée en  $\bar{X} - \bar{Y} + \varepsilon$  et

$$S_p = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}. \quad (2.14)$$

Donc, dans le cas des échantillons indépendants comme dans le cas des échantillons appariés, le test d'hypothèses d'équivalence unilatérale peut être réduit à un test unilatéral ordinaire en transformant légèrement la statistique du test. Les conclusions seront différentes d'un test à l'autre, mais les calculs des régions de rejet et des valeurs critiques sont faits de la même façon.

On conclut avec un exemple numérique d'application de ces tests (Wellek, chapitre 2).

**Exemple 2.1 :** L'étude (Miller, 1990) consiste dans l'analyse de l'effet de la consommation d'olestra (un substitut de la graisse) sur l'absorption des contraceptifs excessivement riches en substances lipophile (*fat substitute olestra on the absorption of highly lipophilic oral contraceptives*). 28 sujets, tous des femmes pré-ménopausées, ont été recrutés pour cette étude. Au cours de 28 jours chaque sujet a consommé 18 mg d'olestra par jour. En même temps, les femmes ont pris des contraceptifs oraux sous forme de combinaison de 300 mg de norgestrel et 30 mg de ethinyl estradiol. Des prélèvements sanguins ont été pris aux jours 12 et 14 et analysés pour la concentration d'ethinyl et d'estradiol. Pour la période où on ingère le placebo, la consommation d'olestra pendant les repas a été remplacée par la consommation des triglycérides, tandis que l'instrument expérimental et celui de mesure sont restés les mêmes que pour la période d'administration du traitement (consommation d'olestra). Dans le tableau 2.2 sont données les valeurs maximales de la concentration du norgestrel (ng/ml) dans le sérum sanguin des 28 femmes. Ces concentrations correspondent au paramètre pharmacocinétique  $C_{\max}$ .

Notons que  $C_{\max}$  représente la concentration maximale du norgestrel (ng/ml) dans le sérum sanguin. Si on considère  $\varepsilon = 1.5$  alors les valeurs  $\tilde{D}_i = D_i + \varepsilon$ , pour  $i = 1, 2, \dots, 28$  et

$\tilde{R}_i^+$  est le rang du  $i$ -ème femme, si on ordonne les valeurs absolues  $|\tilde{D}_i|$  (voir données dans le tableau 2.2).

**Tableau 2.2** Concentrations maximales de norgestrel (ng/ml) dans le sérum sanguin des 28 femmes pour la consommation d'olestra,  $X_i$ , et la consommation des repas contenant du triglycéride ordinaire,  $Y_i$ .

$i$	$X_i$	$Y_i$	$\tilde{D}_i$	$\tilde{R}_i^+$
1	6.03	6.62	0.91	12
2	5.62	6.78	0.34	5
3	6.93	6.85	1.58	18
4	5.86	8.09	-0.73	11
5	8.91	9.18	1.23	15
6	5.86	7.47	-0.11	1
7	9.43	9.90	1.03	13
8	5.30	4.29	2.51	22
9	4.99	3.80	2.69	24
10	6.12	7.01	0.61	8
11	12.45	9.53	4.42	28
12	5.48	6.39	0.59	7
13	6.04	4.63	2.91	25
14	8.32	5.54	4.28	27
15	11.81	11.19	2.12	21
16	8.72	9.55	0.67	9
17	7.01	5.53	2.98	26
18	7.13	6.71	1.92	20
19	6.56	6.53	1.53	17
20	4.22	5.39	0.33	4
21	4.13	4.92	0.71	10
22	6.57	9.92	-1.85	19
23	8.83	10.51	-0.18	2
24	9.05	10.15	0.40	6
25	9.31	9.55	1.26	16
26	7.67	8.95	0.22	3
27	7.66	6.63	2.53	23
28	5.45	8.01	-1.06	14

L'hypothèse initiale du projet était que la consommation d'olestra ne réduisait pas la biodisponibilité du norgestrel (mesuré par  $C_{\max}$ ) de façon significative par rapport à une limite pertinente pré-définie.

Soit  $\theta$  la médiane de la loi des différences intra-sujet  $D_i = X_i - Y_i$ , alors  $\varepsilon = 1.5$  représente la limite pertinente pré-définie. Le test des rangs signés de Wilcoxon (voir section 2.2.2) à été utilisé dans le but de vérifier l'hypothèse du test unilatéral d'équivalence. Les résultats obtenus sont les suivants :

- La somme des rangs  $\tilde{V}_s^+$  des valeurs positives de  $\tilde{D}_i$  est  $\tilde{V}_s^+ = 359$ .
- Pour  $\theta = -\varepsilon = -1.5$ ,  $\tilde{V}_s^+$  suit asymptotiquement une loi normale d'espérance  $E_0(\tilde{V}_s^+)$  et de variance  $Var_0(\tilde{V}_s^+)$ .

La valeur de l'espérance

$$E_0(\tilde{V}_s^+) = \frac{N(N+1)}{4} = 28 * 29 = 203$$

et celle de la variance

$$Var_0(\tilde{V}_s^+) = \frac{N(N+1)(2N+1)}{24} = 1928.5.$$

En tenant compte de la correction de continuité (de 0.5), la valeur approximative de la probabilité observée est égale à

$$p = \Phi\left(\frac{203 - 359 + 0.5}{\sqrt{1928.5}}\right) = \Phi[-3.5410] = 0.0002.$$

D'après la valeur-p obtenue par l'application du test des rangs signés de Wilcoxon, il a été constaté que l'effet d'olestra sur norgestrel était non significatif même pour un seuil de signification de 1%. En d'autres mots, les résultats de l'étude de Miller soutiennent fortement l'hypothèse nulle, que la consommation d'olestra ne réduit pas significativement la biodisponibilité de norgestrel.

## 2.5 Principe d'intervalle de confiance

L'approche la plus populaire et fréquemment utilisée pour les tests unilatéraux et bilatéraux d'équivalence est celle qui utilise l'estimation par intervalles de confiance.

Soit  $\underline{\theta}(\mathcal{X}, \alpha)$  et  $\bar{\theta}(\mathcal{X}, \alpha)$  respectivement les bornes inférieure et supérieure de  $\theta$  au même niveau de confiance unilatéral  $1 - \alpha$ , où  $\mathcal{X}$  représente l'ensemble de toutes les données obtenues pendant l'étude. Par exemple, les valeurs de  $\mathcal{X} = (X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)$ , où  $X_i$  représente les résultats obtenus suite à l'administration du traitement Test,  $T$ , et  $Y_j$  sont les résultats obtenus suite à la l'administration du traitement Référence,  $R$ . Pour un test d'équivalence défini par les hypothèses :

$$H_0 : \theta \leq \theta_0 - \varepsilon_1 \text{ ou } \theta \geq \theta_0 + \varepsilon_2 \quad (2.15)$$

$$\text{contre } H_a : \theta_0 - \varepsilon_1 < \theta < \theta_0 + \varepsilon_2,$$

l'hypothèse nulle (d'inéquivalence) est rejetée si et seulement si l'intervalle aléatoire  $(\underline{\theta}(\mathcal{X}; \alpha), \bar{\theta}(\mathcal{X}; \alpha))$  est complètement couvert par l'intervalle d'équivalence  $(\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2)$  défini par l'hypothèse alternative (d'équivalence). En d'autres mots, l'hypothèse d'équivalence est acceptée si et seulement si les inégalités

$$\underline{\theta}(\mathcal{X}; \alpha) \geq \theta_0 - \varepsilon_1, \quad (2.16)$$

$$\bar{\theta}(\mathcal{X}; \alpha) \leq \theta_0 + \varepsilon_2$$

sont satisfaites simultanément. Sous les hypothèses définies pour les fonctions des paramètres de position, échantillons appariés ou échantillons indépendants (voir section 2.2), le niveau de confiance bilatéral de l'intervalle  $(\underline{\theta}(\mathcal{X}; \alpha), \bar{\theta}(\mathcal{X}; \alpha))$  est  $1 - 2\alpha$ . Cependant, le seuil de signification du test d'équivalence devrait être  $\alpha$  et pas  $2\alpha$ . On va montrer que cela est le cas en suivant le raisonnement suivant.

En effet, par définition, la région de rejet  $C$  pour l'approche par intervalles de confiance peut être présentée sous la forme :

$$C = \{\mathcal{X} | \theta_0 - \varepsilon_1 < \underline{\theta}(\mathcal{X}; \alpha), \bar{\theta}(\mathcal{X}; \alpha) < \theta_0 + \varepsilon_2\}. \quad (2.17)$$

Par définition,

$$P_{\theta} [\theta \in (\underline{\theta}(\mathcal{X}; \alpha), \bar{\theta}(\mathcal{X}; \alpha))] \geq 1 - 2\alpha$$

et, si l'intervalle est symétrique,

$$\begin{cases} P_{\theta}[\theta \leq \underline{\theta}(\mathcal{X}; \alpha)] \leq \alpha, \\ P_{\theta}[\theta \geq \bar{\theta}(\mathcal{X}; \alpha)] \leq \alpha. \end{cases}$$

Donc, si une valeur fixée  $\theta^*$  satisfait

$$\theta^* \leq \theta_0 - \varepsilon_1 < \underline{\theta}(\mathcal{X}; \alpha),$$

on a, en effet,

$$P_{\theta^*}[\underline{\theta}(\mathcal{X}; \alpha) > \theta_0 - \varepsilon_1] \leq P_{\theta^*}[\underline{\theta}(\mathcal{X}; \alpha) \geq \theta^*] \leq \alpha, \quad (2.18)$$

car le domaine  $\{\underline{\theta}(\mathcal{X}; \alpha) > \theta_0 - \varepsilon_1\}$  est inclus dans  $\{\underline{\theta}(\mathcal{X}; \alpha) \geq \theta^*\}$ .

De même, soit une valeur fixée  $\theta^*$  qui satisfait

$$\theta^* \geq \theta_0 + \varepsilon_2 > \bar{\theta}(\mathcal{X}; \alpha).$$

Alors, on a que

$$P_{\theta^*}[\bar{\theta}(\mathcal{X}; \alpha) < \theta_0 + \varepsilon_2] \leq P_{\theta^*}[\bar{\theta}(\mathcal{X}; \alpha) \leq \theta^*] \leq \alpha, \quad (2.19)$$

car le domaine  $\{\bar{\theta}(\mathcal{X}; \alpha) < \theta_0 + \varepsilon_2\}$  est inclus dans  $\{\bar{\theta}(\mathcal{X}; \alpha) \leq \theta^*\}$ .

Par exemple, pour un intervalle de confiance de 90%, la valeur du seuil du test bilatéral est  $\alpha = 5\%$  et pas  $\alpha = 10\%$ , car seulement une des conditions suivantes peut être satisfaite : soit  $\theta^* \leq \theta_0 - \varepsilon_1$ , soit  $\theta^* \geq \theta_0 + \varepsilon_2$ . La probabilité de commettre une erreur de type I est de 5% d'un côté de l'intervalle et 0% de l'autre côté de l'intervalle, ce qui donne en tout 5% et pas 10%.

Donc, même si le niveau de confiance bilatéral de l'intervalle  $(\underline{\theta}(\mathcal{X}; \alpha), \bar{\theta}(\mathcal{X}; \alpha))$  est  $1 - 2\alpha$ , le seuil de signification du test d'équivalence est  $\alpha$  et pas  $2\alpha$ .

## CHAPITRE III

### PLANS D'EXPÉRIENCE EN BIOÉQUIVALENCE

Faire le bon choix du plan d'expérience est essentiel pour une étude de bioéquivalence. Le plan sélectionné joue un rôle important dans l'analyse des données, l'interprétation des résultats et la conclusion par rapport à la bioéquivalence des traitements.

Le choix du plan dépend premièrement de la variabilité entre les observations. Par exemple, la réaction à un certain médicament peut être très différente d'un sujet à l'autre, d'où la variabilité entre les sujets (**variabilité inter-sujet**). Dans ce contexte, un critère de choix du plan d'expérience serait sa capacité d'identifier et d'isoler la variabilité inter-sujet de la **variabilité intra-sujet** (due à la prise des différents traitements par le même sujet) dans l'analyse des données. Le plan d'expérience capable d'éliminer cette variabilité dans la comparaison des traitements serait préféré pour l'analyse des données dans le cadre d'une étude de bioéquivalence. Le plan d'expérience pour ce type d'études est donc principalement choisi en tenant compte des facteurs suivants :

- l'objectif de l'étude;
- le nombre des traitements à être comparé;
- les caractéristiques du traitement et sa pharmacocinétique (absorption, distribution, élimination de l'organisme);
- la disponibilité des sujets;
- la variabilité inter-sujet et la variabilité intra-sujet;
- la durée de l'étude ou le nombre des périodes;
- le coût d'addition d'un sujet comparativement au coût d'addition d'une période;
- le taux d'abandon des sujets.

Les plans d'expérience le plus souvent utilisés dans le cadre d'une étude de bioéquivalence sont les plans complètement aléatoires ou autrement appelés les plans parallèles, les plans d'expérience aux blocs complets, c'est-à-dire les plans croisés et les carrés latins, les plans d'expérience aux blocs incomplets. Tous ces plans peuvent être utiles dépendamment du paramètre à être évalué, des caractéristiques du traitement et/ou des restrictions médicales etc. (Chow et Liu, 2000).

Dans les sections qui suivent, une brève description et quelques exemples sont donnés sur le plan d'expérience parallèle et le plan croisé  $g \times p$ . Plus de détails sont apportées sur le plan d'expérience croisé  $2 \times 2$  à travers des tests classiques et des tests de bioéquivalence.

### 3.1 Plan d'expérience parallèle

Le **plan parallèle** est un plan d'expérience complètement aléatoire, où les sujets ne reçoivent qu'un seul traitement pendant l'étude.

Dans le cadre d'un tel type de plan d'expérience les sujets sont attribués aux traitements de façon aléatoire. Donc, chaque sujet a la même chance de recevoir le traitement qui lui sera administré au long de l'expérience. Normalement, le nombre des sujets attribué aux différents groupes des traitements est le même.

L'exemple le plus simple d'un plan parallèle est celui qui compare deux traitements. Soient deux traitements administrés à vingt sujets de façon aléatoire. Dix d'entre eux reçoivent le nouveau traitement, Test, et les autres dix sujets reçoivent le traitement courant ou standard, Référence, de la façon suivante:

**Tableau 3.1** Exemple de plan parallèle

Traitement	Sujet
Test	2, 3, 5, 9, 10, 13, 15, 17, 18, 20
Référence	1, 4, 6, 7, 8, 11, 12, 14, 16, 19

Le plan parallèle peut être décrit par le modèle linéaire:

$$y_{ij} = \mu + \tau_j + e_{ij},$$

où  $y$  représente la variable dépendante et  $y_{ij}$  est l'observation obtenue pour le sujet  $i$  dans le groupe du traitement  $j$ , pour tout  $i = 1, 2, \dots, n_t$  et  $j = 1, 2, \dots, t$ , et :

- $\mu$  est la moyenne globale de la population,  $\mu = \bar{\mu} = \sum_{j=1}^t (\mu_j / t)$ , où  $\mu_j$  représente la moyenne de la population du traitement  $j$  ;
- $\tau_j$  représente l'effet direct du traitement  $j$ , c'est-à-dire la déviation de la moyenne  $\mu_j$  par rapport à la moyenne globale  $\mu$ ,  $\tau_j = \mu_j - \mu$  et  $\sum_{j=1}^t \tau_j = 0$  ;
- $e_{ij}$  est l'erreur expérimentale de l'observation  $y_{ij}$ ,  $e_{ij} = y_{ij} - \mu_j$  ;
- les erreurs  $\{e_{ij}\}$  sont des observations indépendantes de moyennes  $E(e_{ij}) = 0$  et de variance  $V(e_{ij}) = \sigma_e^2$  supposée égale pour toutes les populations des traitements  $j$ .

Pour tester l'hypothèse nulle :

$$H_0: \tau_1 = \tau_2 = \dots = \tau_t$$

par rapport à

$$H_a: \text{il existe au moins un } j \neq k \text{ ou } \tau_j \neq \tau_k$$

la statistique suivante est utilisée :

$$F_0 = \frac{MST}{MSE}$$

où

- $F_0$  suit une loi  $F$  sous l'hypothèse nulle;
- $MST$  décrit la variation entre les traitements;
- $MSE$  décrit l'erreur expérimentale.

Les formules de calcul de  $MST$  et de  $MSE$  sont données dans le tableau d'analyse de variance (tableau 3.2).

**Tableau 3.2** Analyse de Variance pour un plan d'expérience parallèle

Source de variation	Degrés de libertés	Somme des carrés (SS)	Moyenne des carrés
Traitements	$t - 1$	$SST = \sum_{j=1}^t n_i (\bar{y}_j - \bar{y}_{..})^2$	$MST = \frac{SST}{t - 1}$
Erreur	$N - t$	$SSE = \sum_{j=1}^t \sum_{i=1}^{n_i} (y_{ij} - \bar{y}_{.j})^2$	$MSE = \frac{SSE}{N - t}$
Totale	$N - 1$	$SSTotal = \sum_{j=1}^t \sum_{i=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$	

L'hypothèse  $H_0$  est rejetée au seuil de signification  $\alpha$ , si et seulement si

$$F_0 > F_{\alpha, (t-1), (N-t)},$$

où  $N = \sum_{i=1}^t n_i$  représente le nombre total des sujets participants à l'étude et  $F_{\alpha, (t-1), (N-t)}$  est le  $\alpha$ -quantile de la loi  $F$  avec  $t - 1$  et  $N - t$  degrés de liberté.

Il faut noter que le plan d'expérience parallèle est peu utilisé dans les études de bioéquivalence. Particulièrement, cela s'explique par le fait que la variabilité intra-sujet, qui joue un rôle important dans l'évaluation de la bioéquivalence, ne peut pas être identifiée et calculée séparément de la variabilité inter-sujet. Ces deux sources de variation ne peuvent pas

être isolées, car les sujets reçoivent la même formulation tout au long de l'expérience. Cela implique qu'en utilisant le plan parallèle la bioéquivalence est établie par des comparaisons basées sur les variabilités intra- et inter-sujet. Malgré le fait que le plan parallèle ne soit pas capable d'isoler et d'éliminer la variabilité inter-sujet de la comparaison des formulations, ce plan est considéré comme une alternative aux autres types de plans d'expérience utilisés dans le cadre des études de bioéquivalence dans les conditions suivantes:

- le traitement est potentiellement toxique et/ou d'une très longue durée d'élimination;
- la population d'intérêt est composée de patients très malades;
- le coût de l'ajout d'un groupe supplémentaire de sujets est plus petit comparativement au coût de l'ajout d'une autre période expérimentale.

### 3.2 Plan croisé $g \times p$

Le **plan d'expérience croisé** représente une version modifiée du plan aléatoire par blocs. Généralement, le **bloc** peut être composé d'une unité ou d'un group d'unités expérimentales homogènes. On peut considérer comme critère d'homogénéité l'âge ou le poids du sujet, par exemple, si l'unité expérimentale représente la personne. Si le bon critère est choisi, la variation entre les unités expérimentales du même bloc est plus petite que la variation entre les unités expérimentales des différents blocs.

Dans le cadre d'un plan d'expérience croisé, le bloc est composé d'une seule unité expérimentale. La même unité expérimentale reçoit plusieurs traitements pendant l'étude. Ces traitements sont administrés pendant différentes périodes de temps nommées **périodes expérimentales**. Quand les traitements sont comparés sur la même unité expérimentale (le même sujet), la variation entre les unités expérimentales est isolée et éliminée de l'erreur expérimentale. Les unités expérimentales individuelles sont utilisées comme bloc dans le cadre des plans croisés pour diminuer l'erreur expérimentale et pour augmenter l'efficacité de l'étude. Comparativement aux plans complètement aléatoires, la comparaison des traitements d'un plan aux blocs est plus précise, car la variation bloc (entre les blocs) est éliminée de la comparaison des traitements sur la même unité expérimentale. Donc, c'est la précision plus élevée qui constitue l'avantage principal des plans par blocs. Un bloc est considéré *complet*,

si toutes les unités expérimentales du bloc reçoivent tous les traitements à être comparés pendant l'étude. L'ordre dans lequel les unités expérimentales reçoivent ces traitements est nommé **séquence**. Un plan d'expérience croisé est composé de plusieurs séquences, définies de telle façon pour que l'effet traitement ne soit pas confondu avec l'effet période; plus précisément, pour que ces deux effets puissent être distingués l'un de l'autre. Par exemple, pour le plan croisé  $2 \times 2$  (composé des deux séquences et deux périodes), utilisé pour comparer deux traitements  $A$  et  $B$ , dans la première séquence le traitement  $A$  est administré dans la période 1 et le traitement  $B$  est administré dans la période 2. L'ordre d'administration des traitements change dans la deuxième séquence, où le traitement  $B$  est administré dans la période 1 et le traitement  $A$  est administré dans la période 2. Cette façon de définir les séquences, permet d'éviter que les effets période et traitement soient confondus. Le plan croisé, dont le nombre de séquences est égal à  $g$  et le nombre de périodes expérimentales est égal à  $p$ , est nommé **plan croisé  $g \times p$** .

Généralement, dans le cadre des études de bioéquivalence les unités expérimentales ou bien les blocs représentent des sujets non malades. Chaque sujet est attribué aléatoirement à une des séquences. Donc les séquences sont composées des réponses de sujets différents.

Pour le plan croisé on distingue :

- l'effet **direct du traitement** qui représente l'effet du traitement mesuré pendant la période expérimentale;
- l'effet **résidu** (*carry-over*) qui représente l'effet du traitement qui dure encore après la fin de la période expérimentale.

Si l'effet résidu dure pendant une seule période, il s'appelle effet de **premier rang**. Dans le cas où il dure pendant jusqu'à  $c$  périodes, il s'appelle effet de **rang  $c$** .

Généralement, dans le cadre des études de bioéquivalence, les périodes expérimentales sont séparées par un intervalle de temps nommé **washout** (lavage). Pendant cette période, l'effet du premier traitement devrait être éliminé de l'organisme pour ne pas être présent dans la période suivante. Cet intervalle doit être assez long pour amener le sujet à l'état physiologique ou psychologique initial. Cette période est définie, d'habitude, comme cinq

fois la demi-vie du traitement ( $5t_{1/2}$ ). Il faut rappeler que le paramètre  $t_{1/2}$  correspond au temps nécessaire pour que, après l'administration du traitement, sa concentration diminue de moitié. D'après la littérature, une période de temps égale à cinq demi-vies est considérée suffisante pour que 97% de l'effet du traitement soit éliminé de l'organisme.

Notons que, même si le *washout* est assez long, l'état physiologique peut être suffisamment affecté pour qu'un certain effet du traitement soit encore présent dans les périodes expérimentales suivantes. En utilisant un plan croisé dans ce cas, une période d'élimination suffisamment longue doit être définie pour que le traitement soit éliminé complètement de l'organisme. L'expérience risque de durer trop long temps, ce qui peut amener à un grand nombre d'abandons de l'étude. Cela constitue un des inconvénients majeurs des plans d'expérience croisés.

### 3.2.1 Exemples de plans croisés

En ce qui suit, quelques exemples des plans d'expérience croisés sont présentés.

#### 1. Plan d'expérience 2 x 2

Dans le cadre d'une étude de bioéquivalence un nouveau traitement, appelé Test (T), est comparé à un traitement standard, appelé Référence (R), en utilisant le plan croisé suivant :

**Tableau 3.3** Plan croisé 2 x 2

Séquence	Période	
	1	2
1	Test (T)	Référence (R)
2	Référence (R)	Test (T)

Ce dernier représente un plan à 2 séquences et 2 périodes. Chaque sujet reçoit les traitements Test et Référence dans deux périodes expérimentales différentes. Pour éviter le biais, les sujets sont assignés de façon aléatoire à une de ces deux séquences, TR et RT. Les périodes expérimentales sont séparées par une période de *washout* assez longue pour que l'effet résidu soit complètement éliminé. Le plan d'expérience 2 x 2 représente un carré latin. Le **carré latin** est un plan d'expérience dont les  $L$  traitements sont arrangés dans un carré de  $L$  lignes et  $L$  colonnes où chaque ligne et chaque colonne ne contient qu'un seul traitement, et  $L = 2$  dans ce cas.

## 2. Plans d'ordre supérieur

Deux traitements ne sont pas toujours comparés en utilisant un plan croisé 2 x 2. Dans certaines situations, des plans croisés d'ordre supérieur, dont le nombre des périodes ou des séquences n'est pas nécessairement égal au nombre des traitements, sont plus appropriés. Typiquement, on peut avoir un traitement qui est **hautement variable**, c'est-à-dire il y a une grande dispersion intra-sujet (une déviation standard de plus de 0.30, Patterson et Jones (2006), chapitre 2), ou un traitement dont l'effet résidu dure longtemps dû à une longue demi-vie, à un faible métabolisme ou à d'autres facteurs qui influence le processus d'élimination plus rapide du traitement de l'organisme. Dans ces cas, deux traitements sont comparés en utilisant les plans croisés 2 x 3 ou 2 x 4. Pour ces plans croisés, le nombre de périodes est plus grand que le nombre de traitements.

D'après le plan d'expérience 2 x 3 (ou **demi-répétés**), les sujets vont recevoir les deux traitements. Cependant, un traitement est administré deux fois. Lequel des deux traitements est administré deux fois dépend de la séquence qu'on a attribué aléatoirement au sujet. Il existe six séquences possibles pour ce plan d'expérience, mais juste deux d'entre elles sont le plus souvent utilisées (indiquées en caractères gras dans le tableau 3.4).

**Tableau 3.4** Séquences du plan croisé 2 x 3

Séquence	Période		
	1	2	3
1	<b>T</b>	<b>R</b>	<b>R</b>
2	<b>R</b>	<b>T</b>	<b>T</b>
3	T	T	R
4	R	R	T
5	T	R	T
6	R	T	R

D'après le plan d'expérience 2 x 4 (**pleinement répété**), les sujets reçoivent deux fois chaque traitement. Il existe six séquences possibles pour ce plan d'expérience. En pratique, deux d'entre elles sont plus souvent utilisées (indiquées en caractères gras dans le tableau 3.5).

**Tableau 3.5** Séquences du plan croisé 2 x 4

Séquence	Période			
	1	2	3	4
1	T	R	R	T
2	R	T	T	R
3	T	T	R	R
4	R	R	T	T
5	<b>T</b>	<b>R</b>	<b>T</b>	<b>R</b>
6	<b>R</b>	<b>T</b>	<b>R</b>	<b>T</b>

### 3. Plan d'expérience 3 x 3

Dans le cadre d'une étude, c'est possible que des plans croisés pour comparer plus de deux traitements soient utilisés pour tester la bioéquivalence. Cependant, ils sont moins courants que les plans déjà mentionnés.

Supposons que, deux formulations référence R et S sont comparées à la formulation test T. La formulation R peut représenter une dose composée de trois comprimés de 100 mg et la formulation S peut représenter une dose composée d'un comprimé de 200 mg et d'un comprimé de 100 mg, tandis que la formulation test peut être composée d'un seul comprimé de 300 mg. Le but de l'étude est de comparer la bioéquivalence de T avec R et la bioéquivalence de T avec S. Le plan d'expérience 3 x 3 (voir tableau 3.6) est à utiliser, où les sujets sont attribués aléatoirement à une des six séquences possibles.

**Tableau 3.6** Séquences du plan croisé 3 x 3

Séquence	Période		
	1	2	3
1	T	R	S
2	R	S	T
3	S	T	R
4	T	S	R
5	R	T	S
6	S	R	T

On constate que le plan 3 x 3 est formé de deux carré latins, où l'un d'entre eux est formé par les séquences 1, 2, 3 et l'autre est formé par les séquences 4, 5, 6. En pratique, un seul carré latin est souvent utilisé pour comparer les traitements. Il est préférable, cependant, que le plan complet soit utilisé.

#### 4. Plan d'expérience 4 x 4

Si jamais la bioéquivalence de deux formulations test et deux formulations référence est à tester, le plan d'expérience 4 x 4, composés des séquences suivantes, est préférable.

**Tableau 3.7** Séquences du plan croisé 4 x 4

Séquence	Période			
	1	2	3	4
1	A	D	B	C
2	B	A	C	D
3	C	B	D	A
4	D	C	A	B

On a noté par  $A, C$  la dose élevée et la dose basse du traitement Référence, respectivement, et par  $B, D$  la dose élevée et la dose basse du traitement Test, respectivement.

#### 3.2.2 Modèle linéaire général

Supposons que pour comparer  $L$  traitements ( $t = 1, 2, \dots, L$ ) on utilise le plan croisé  $g \times p$ . Comme c'était déjà mentionné, ce plan est composé de  $p$  périodes ( $j = 1, 2, \dots, p$ ) et de  $g$  séquences ( $k = 1, 2, \dots, g$ ). Le nombre de périodes expérimentales n'est pas nécessairement égal au nombre de traitements. Un total de  $n_k$  sujets ( $i = 1, 2, \dots, n_k$ ) est attribué aléatoirement à chacune de ces séquences.

Soit  $y_{ijk}$  l'observation (par exemple la valeur observée de  $AUC$  ou  $C_{max}$ ) du sujet  $i$  de la séquence  $k$  dans la période  $j$ .

*Définition 3.1 :* Un facteur  $A_1$  à  $n$  niveaux  $\{a_{11}, a_{12}, \dots, a_{1n}\}$  est dit **emboîté** dans un facteur  $A_2$  à  $m$  niveaux  $\{a_{21}, a_{22}, \dots, a_{2m}\}$ , lorsque à chaque niveau  $a_1 \in \{a_{11}, a_{12}, \dots, a_{1n}\}$  correspond un seul niveau  $a_2 \in \{a_{21}, a_{22}, \dots, a_{2m}\}$ .

Le modèle linéaire qui correspond à ce plan croisé peut être décrit par la formule :

$$y_{ijk} = \mu + \gamma_k + S_{ik} + \pi_j + \tau_{(j,k)} + \lambda_{(j-1,k)} + e_{ijk} \quad (3.1)$$

$$i = 1, 2, \dots, n_k, j = 1, 2, \dots, p, k = 1, 2, \dots, g$$

où :

- $\mu$  est la moyenne globale;
- $\gamma_k$  représente l'effet fixe de la séquence  $k$ ,  $k = 1, 2, \dots, g$ , et on a  $\sum_{k=1}^g \gamma_k = 0$ ;
- $S_{ik}$  est l'effet aléatoire du sujet  $i$  emboîté dans la séquence  $k$  ( $i = 1, 2, \dots, n_k$ ), c'est-à-dire que chaque sujet n'est attribué qu'à une des deux séquences et pas aux deux séquences simultanément;
- $\pi_j$  est l'effet fixe de la période  $j$ ,  $j = 1, 2, \dots, p$ , et on a  $\sum_{j=1}^p \pi_j = 0$ ;
- $\tau_{(j,k)}$  représente l'effet direct du traitement administré dans la période  $j$  de la séquence  $k$ . C'est un effet fixe pour lequel  $\sum_{k=1}^g \tau_{(j,k)} = 0$ ;
- $\lambda_{(j-1,k)}$  est l'effet fixe du résidu (effet résidu de premier rang) du traitement administré dans la période  $j - 1$  de la séquence  $k$ , où  $\lambda_{(0,k)} = 0$  et  $\sum_{k=1}^g \lambda_{(j-1,k)} = 0$ ;
- $e_{ijk}$  est l'erreur expérimentale pour l'observation  $y_{ijk}$ .

On a les suppositions suivantes:

1. les effets aléatoires  $\{S_{ik}\}$  sont indépendants et identiquement distribués de moyenne  $E(S_{ik}) = 0$  et de variance  $V(S_{ik}) = \sigma_B^2$ ,  $i = 1, 2, \dots, n_k$ ,  $k = 1, 2, \dots, g$ ;
2. les erreurs  $\{e_{ijk}\}$ , sont indépendantes de moyenne  $E(e_{ijk}) = 0$  et de variance  $V(e_{ijk}) = \sigma_e^2$ ,  $i = 1, 2, \dots, n_k$ ,  $j = 1, 2, \dots, p$ ,  $k = 1, 2, \dots, g$ ,  $t = 1, 2, \dots, L$ , où  $L$  représente le nombre de traitements à comparer. Par exemple, dans le plan croisé  $2 \times 2$  le nombre de traitements  $L = 2$ , dans les plans d'ordre supérieur  $2 \times 3$  ou  $2 \times 4$  le nombre de traitements  $L = 2$ , mais dans le plan croisé  $3 \times 3$  le nombre de traitements  $L = 3$  etc.;
3. les ensembles  $\{S_{ik}\}$  et  $\{e_{ijk}\}$  sont mutuellement indépendants.

La variance  $\sigma_B^2$  décrit la variabilité inter-sujet, tandis que la variance  $\sigma_t^2$  décrit la variabilité intra-sujet pour le traitement  $t = 1, 2, \dots, L$ .

Soit  $\bar{y}_{.1k}, \bar{y}_{.2k}, \dots, \bar{y}_{.pk}$  les moyennes observées pour les périodes  $1, 2, \dots, p$  de la séquence  $k$  obtenues par la formule  $\bar{y}_{.jk} = (1/n_k) \sum_{i=1}^{n_k} y_{ijk}$ ,  $j = 1, 2, \dots, p$  et  $k = 1, 2, \dots, g$ .

Sous l'hypothèse de la normalité de l'effet aléatoire et de l'erreur expérimentale, l'effet période, l'effet direct du traitement et l'effet résidu du modèle peuvent être estimés en utilisant des contrastes linéaires appropriés définies à partir de ces  $g * p$  moyennes.

*Définition 3.2 :* Le **contraste linéaire intra-sujet** de la séquence  $k$  représente une combinaison linéaire des moyennes observées  $\bar{y}_{.1k}, \bar{y}_{.2k}, \dots, \bar{y}_{.pk}$ , c'est-à-dire :

$$l = c_1 \bar{y}_{.1k} + c_2 \bar{y}_{.2k} + \dots + c_p \bar{y}_{.pk},$$

où  $\sum_{j=1}^p c_j = 0$ .

On constate que la variance du contraste  $l$  ne consiste que de la variabilité intra-sujet  $\sigma_t^2$ ,  $t = 1, 2, \dots, L$ . Alors, les inférences statistiques pour l'effet période, l'effet direct du traitement et l'effet résidu peuvent être réalisées en se basant sur la variabilité intra-sujet. Dans la section suivante des tests classiques sont appliqués afin de réaliser les inférences statistiques sur les effets fixes du plan croisé  $2 \times 2$ .

### 3.3 Inférence statistique pour le plan croisé $2 \times 2$ (tests classiques)

Le plan le plus populaire dans le cadre des études de bioéquivalence d'un nouveau traitement, Test (T), et un traitement standard, Référence (R), est le plan croisé  $2 \times 2$ . Tel que vu auparavant, ce plan est composé de deux séquences et deux périodes expérimentales. Chaque sujet est attribué aléatoirement à l'une des séquences RT ou TR. Les sujets reçoivent les deux traitements durant l'étude, un dans la période expérimentale un et l'autre dans la période expérimentale deux. Typiquement, les périodes expérimentales sont séparées par une période *washout* assez longue pour que l'effet résidu d'un traitement ne soit pas rapporté d'une période expérimentale à l'autre.

Soit  $y_{ijk}$  la valeur observée (par exemple  $AUC$  ou  $C_{max}$ ) pour le sujet  $i$  ( $i = 1, 2, \dots, n_k$ ) de la séquence  $k$  ( $k = 1, 2$ ) dans la période  $j$  ( $j = 1, 2$ ).

D'après Chow et Liu, 2000, chapitre 3, le modèle linéaire du plan croisé 2 x 2 peut être décrit par les quatre combinaisons période-séquence suivantes (voir tableau 3.8):

**Tableau 3.8** Combinaisons période-séquence du modèle linéaire du plan croisé 2x2

Séquence	Période	
	1	2
1 (RT)	$y_{i11} = \mu + S_{i1} + \pi_1 + \tau_R + e_{i11}$	$y_{i21} = \mu + S_{i1} + \pi_2 + \tau_T + \lambda_R + e_{i21}$
2 (TR)	$y_{i'12} = \mu + S_{i'2} + \pi_1 + \tau_T + e_{i'12}$	$y_{i'22} = \mu + S_{i'2} + \pi_2 + \tau_R + \lambda_T + e_{i'22}$

où :

- $\mu$  est la moyenne globale;
- $S_{i1}$  et  $S_{i'2}$  sont les effets aléatoires des sujets  $i$  et  $i'$  emboîtés dans la séquence 1 et 2, respectivement,  $i = 1, 2, \dots, n_1$  et  $i' = 1, 2, \dots, n_2$ ;
- $\pi_1$  et  $\pi_2$  sont les effets fixes de la période 1 et 2, respectivement, et  $\pi_1 + \pi_2 = 0$ ;
- $\tau_T$  et  $\tau_R$  représentent les effets directs du traitement T et du traitement R, respectivement, et  $\tau_T + \tau_R = 0$ ;
- $\lambda_T$  et  $\lambda_R$  représentent les effets résidu dû au traitement T ou au traitement R, respectivement, et  $\lambda_T + \lambda_R = 0$ . Il n'existe pas d'effet résidu dans la première période, c'est-à-dire qu'il se peut qu'un effet résidu soit présent que pour la deuxième période;
- $e_{i11}$ ,  $e_{i21}$ ,  $e_{i'12}$  et  $e_{i'22}$  sont les erreurs expérimentales pour les observations  $y_{i11}$ ,  $y_{i21}$ ,  $y_{i'12}$  et  $y_{i'22}$ , respectivement,  $i = 1, 2, \dots, n_1$  et  $i' = 1, 2, \dots, n_2$ .

On a les suppositions suivantes:

1. les effets aléatoires  $S_{i1}$  et  $S_{i'2}$ , sont indépendants, identiquement distribués et ils suivent une loi normale  $N(0, \sigma_b^2)$ ;
2. les erreurs expérimentales  $\{e_{ijk}\}$  ( $i = 1, 2, \dots, n_k, j = 1, 2, k = 1, 2$ ) sont indépendantes, identiquement distribuées et elles suivent une loi normale  $N(0, \sigma_w^2)$ ;
3. les ensembles  $\{S_{ik}\}$  et  $\{e_{ijk}\}$  ( $i = 1, 2, \dots, n_k, j = 1, 2, k = 1, 2$ ) sont mutuellement indépendants.

Les hypothèses du modèle défini par les effets du tableau 3.8 sont plus restreintes, comparativement aux hypothèses du modèle du plan croisé  $g \times p$  à  $g$  séquences et  $p$  périodes défini dans la section précédente, où les ensembles  $\{S_{ik}\}$ , et  $\{e_{ijk}\}$  ( $i = 1, 2, \dots, n_k, j = 1, 2, \dots, p, k = 1, 2, \dots, g$ ) ne suivaient pas nécessairement des lois normales. De plus, les variabilités intra-sujet de  $L$  traitements  $\sigma_t^2$  ( $t = 1, 2, \dots, L$ ), variait d'un traitement à l'autre pour le plan croisé  $g \times p$ . Dans le cas du plan  $2 \times 2$ , la variabilité intra-sujet  $\sigma_w^2$  est la même pour les deux traitements.

Notons que dans notre cas  $L = 2$ .

Les effets fixes du plan croisé  $2 \times 2$  pour chacune des quatre combinaisons période-séquence, sont présentés dans le tableau suivant:

**Tableau 3.9** Les effets fixes du plan croisé  $2 \times 2$  sans effet séquence

Séquence	Période	
	1	2
1 (RT)	$\mu_{11} = E(y_{i11}) = \mu + \pi_1 + \tau_R$	$\mu_{21} = E(y_{i21}) = \mu + \pi_2 + \tau_T + \lambda_R$
2 (TR)	$\mu_{12} = E(y_{i12}) = \mu + \pi_1 + \tau_T$	$\mu_{22} = E(y_{i22}) = \mu + \pi_2 + \tau_R + \lambda_T$

où on a supposé que  $\pi_1 + \pi_2 = 0$ ,  $\tau_R + \tau_T = 0$  et  $\lambda_R + \lambda_T = 0$ .

L'objectif est d'isoler l'effet direct du traitement des autres effets et de l'estimer.

Il faut noter que le modèle linéaire définit d'après Chow et Liu, 2000, chapitre 3, ne contient que quatre paramètres à estimer, c'est-à-dire  $\mu$ ,  $\pi$ ,  $\tau$  et  $\lambda$ . D'après Patterson et Jonce, 2006, chapitre 3, étant donné que les sujets sont attribués aléatoirement aux séquences, l'effet séquence peut être supposé absent. Si jamais l'effet séquences était conservé dans le modèle, on n'aurait pas eu la possibilité d'estimer tous les cinq paramètres, car juste quatre moyennes observées sont disponibles (voir tableau 3.10). Dans ce cas l'effet résidu est confondu avec l'effet séquence (Chow et Liu, chapitre 2).

**Tableau 3.10** Les effets fixes du plan croisé 2 x 2 avec effet séquence

Séquence	Période	
	1	2
1 (RT)	$\mu_{11} = E(y_{i11}) = \mu + \gamma_1 + \pi_1 + \tau_R$	$\mu_{21} = E(y_{i21}) = \mu + \gamma_1 - \pi_1 - \tau_R + \lambda_R$
2 (TR)	$\mu_{12} = E(y_{i12}) = \mu - \gamma_1 + \pi_1 - \tau_R$	$\mu_{22} = E(y_{i22}) = \mu - \gamma_1 - \pi_1 + \tau_R - \lambda_R$

Les espérances du tableau 3.10 ont été obtenues en substituant les égalités suivantes  $\tau_T = -\tau_R$ ,  $\lambda_T = -\lambda_R$ ,  $\gamma_2 = -\gamma_1$  et  $\pi_2 = -\pi_1$  dans le tableau 3.9. Ces égalités ont été obtenues à partir des suppositions  $\tau_R + \tau_T = 0$ ,  $\lambda_R + \lambda_T = 0$ ,  $\gamma_1 + \gamma_2 = 0$ ,  $\pi_1 + \pi_2 = 0$ , respectivement.

Les effets aléatoires du plan croisé 2 x 2, pour chacune des quatre combinaisons période-séquence, sont présentés dans le tableau 3.11 suivant:

**Tableau 3.11** Les effets aléatoires du plan croisé 2 x 2

Séquence	Période	
	1	2
1 (RT)	$S_{i1} + e_{i11}$	$S_{i1} + e_{i21}$
2 (TR)	$S_{i'2} + e_{i'12}$	$S_{i'2} + e_{i'22}$

$i = 1, 2, \dots, n_1$  et  $i' = 1, 2, \dots, n_2$ .

La variance du sujet  $i$  de la séquence  $k$  dans la période  $j$ , est

$$V(y_{ijk}) = V(S_{ik} + e_{ijk}) = \sigma_b^2 + \sigma_w^2. \quad (3.2)$$

On retrouve aussi la formule de la covariance entre deux observations dans la même séquence, mais dans des périodes différentes, qui est égale à :

$$\begin{aligned} Cov(y_{i1k}, y_{i2k}) &= Cov(S_{ik} + e_{i1k}, S_{ik} + e_{i2k}) = Cov(S_{ik}, S_{ik}) = \\ V(S_{ik}) &= \sigma_b^2, \quad i = 1, 2, \dots, n_k, \quad k = 1, 2. \end{aligned} \quad (3.3)$$

La corrélation,  $\rho$ , entre deux observations du même sujet sur des périodes différentes se calcule selon la formule :

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}.$$

Notons que les formules (3.3.2) et (3.3.3) restent valables si les lois des erreurs expérimentales ne sont pas normales.

Sous les hypothèses du modèle définie dans le tableau 3.8, des inférences statistiques peuvent être réalisées pour les effets fixes en utilisant des tests  $t$  de Student pour deux échantillons indépendants. Plus de détails sur ces inférences seront donnés dans les sections suivantes.

### 3.3.1 L'effet résidu

Avant de tester l'effet direct du traitement on va se rassurer que la présence de l'effet résidu n'est pas significative, car, au cas contraire, il n'existe pas d'estimateur non-biaisé pour l'effet direct du traitement.

Soit la somme entre l'observation de la période 1 et l'observation de la période 2 du sujet  $i$  ( $i = 1, 2, \dots, n_k$ ) de la séquence  $k$  ( $k = 1, 2$ ) notée par  $u_{ik}$ . Alors, pour chaque sujet on peut considérer l'égalité suivante :

$$u_{ik} = y_{i1k} + y_{i2k}, \quad i = 1, 2, \dots, n_k, \quad k = 1, 2. \quad (3.4)$$

Les espérances mathématiques des variables  $u_{ik}$  ( $i = 1, 2, \dots, n_k, k = 1, 2$ ) sont données dans le tableau suivant :

**Tableau 3.12** Les espérances mathématiques des variables  $u_{ik}$

Séquence	
1	$E(u_{i1}) = E(y_{i11} + y_{i21}) = 2\mu + (\pi_1 + \pi_2) + (\tau_R + \tau_T) + \lambda_R = 2\mu + \lambda_R$
2	$E(u_{i2}) = E(y_{i12} + y_{i22}) = 2\mu + (\pi_1 + \pi_2) + (\tau_T + \tau_R) + \lambda_T = 2\mu + \lambda_T$

où  $E(S_{ik}) = 0$ ,  $E(e_{ijk}) = 0$  ( $i = 1, 2, \dots, n_k, k = 1, 2, j = 1, 2$ ) et  $\pi_1 + \pi_2 = 0$ ,  $\tau_T + \tau_R = 0$ .

La variance des  $u_{ik}$  ( $i = 1, 2, \dots, n_k, k = 1, 2$ ) est décrite par la formule:

$$V(u_{ik}) = V(y_{i1k} + y_{i2k}) = V(y_{i1k}) + V(y_{i2k}) + 2Cov(y_{i1k}, y_{i2k}) = 2(\sigma_w^2 + \sigma_b^2) + 2\sigma_b^2 = 2(\sigma_w^2 + 2\sigma_b^2).$$

Donc on a

$$\sigma_u^2 = V(u_{ik}) = 2(\sigma_w^2 + 2\sigma_b^2). \quad (3.5)$$

Pour tester l'effet résidu, on va considérer le paramètre  $\lambda = \lambda_T - \lambda_R$ . Sous la contrainte  $\lambda_T + \lambda_R = 0$ , les effets résidus des deux traitements sont égaux si et seulement si  $\lambda_T = \lambda_R = 0$ . Alors, tester l'absence de l'effet résidu c'est la même chose que montrer l'égalité entre l'effet résidu du traitement Test et l'effet résidu du traitement Référence,  $\lambda_T = \lambda_R$ . Dans le cas où il n'existe pas d'effet résidu, l'effet direct du traitement est estimé à partir des observations des

deux périodes. Si jamais les effets résidu des deux traitements ne sont pas égaux, il n'existe pas d'estimateur non-biaisé pour estimer l'effet direct du traitement, d'où l'importance de tester la signification de l'effet résidu.

On définit les hypothèses suivantes :

$$H_0: \lambda = 0 \text{ (ou } \lambda_T = \lambda_R) \quad (3.6)$$

contre

$$H_a: \lambda \neq 0 \text{ (ou } \lambda_T \neq \lambda_R).$$

Si jamais l'hypothèse  $H_0$  est rejetée, cela signifie que l'effet résidu est significatif.

Notons par

$$\bar{u}_{.k} = \frac{1}{n_k} \sum_{i=1}^{n_k} u_{ik}, \quad k = 1, 2, \quad (3.7)$$

les moyennes des valeurs  $u_{ik}$  pour chacune des séquences. Alors,  $\bar{u}_{.1}$  et  $\bar{u}_{.2}$  représentent les moyennes de deux échantillons aléatoires indépendants tirés de populations qui suivent des lois normales de la même variance. Dans ces conditions, l'inférence statistique sur  $\lambda$  peut se faire facilement en appliquant un test t de Student pour deux échantillons indépendants. Étant donné que  $\lambda$  est inconnu, il sera estimé par le contraste linéaire des moyennes  $\bar{u}_{.1}$  et  $\bar{u}_{.2}$  :

$$\hat{\lambda} = \bar{u}_{.2} - \bar{u}_{.1} = (\bar{y}_{.12} + \bar{y}_{.22}) - (\bar{y}_{.11} + \bar{y}_{.21}). \quad (3.8)$$

Sous les hypothèses du modèle du plan croisé 2 x 2 (voir tableau 3.8),  $\hat{\lambda}$  suit une loi normale d'espérance mathématique suivante :

$$E(\hat{\lambda}) = E(\bar{u}_{.2} - \bar{u}_{.1}) = E(\bar{u}_{.2}) - E(\bar{u}_{.1}) = (2\mu + \lambda_T) - (2\mu + \lambda_R) = \lambda_T - \lambda_R = \lambda,$$

c'est-à-dire

$$E(\hat{\lambda}) = \lambda, \quad (3.9)$$

où

$$E(\bar{u}_k) = \frac{1}{n_k} \sum_{i=1}^{n_k} E(u_{ik}) = \begin{cases} \frac{n_1}{n_1} (2\mu + \lambda_R) = 2\mu + \lambda_R \\ \frac{n_2}{n_2} (2\mu + \lambda_T) = 2\mu + \lambda_T \end{cases}.$$

La variance est :

$$V(\hat{\lambda}) = 2(2\sigma_b^2 + \sigma_w^2) \left( \frac{1}{n_1} + \frac{1}{n_2} \right) = \sigma_u^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right), \quad (3.10)$$

et elle s'obtient de la façon suivante :

$$\begin{aligned} V(\bar{u}_k) &= V\left(\frac{1}{n_k} \sum_{i=1}^{n_k} u_{ik}\right) = \frac{1}{n_k^2} V\left(\sum_{i=1}^{n_k} u_{ik}\right) \\ &= \frac{1}{n_k^2} \sum_{i=1}^{n_k} V(u_{ik}) = \frac{n_k}{n_k^2} (2(\sigma_w^2 + 2\sigma_b^2)) = \frac{2}{n_k} (\sigma_w^2 + 2\sigma_b^2), \end{aligned}$$

car  $u_{ik}$  sont indépendants pour  $i = 1, 2, \dots, n_k$ . En particulier,  $Cov(\bar{u}_2, \bar{u}_1) = 0$  par l'indépendance entre les séquences.

La variance  $V(\hat{\lambda})$  s'estime par  $\hat{V}(\hat{\lambda}) = \hat{\sigma}_u^2 (1/n_1 + 1/n_2)$ , où  $\sigma_u^2$  est remplacé par  $\hat{\sigma}_u^2$ , où :

$$\hat{\sigma}_u^2 = \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^2 \sum_{i=1}^{n_k} (u_{ik} - \bar{u}_k)^2. \quad (3.11)$$

L'estimateur  $\hat{\lambda}$  est un estimateur non-biaisé de  $\lambda$  et  $\hat{\sigma}_u^2$  est un estimateur non-biaisé de  $\sigma_u^2$ . De plus,  $(n_1 + n_2 - 2)\hat{\sigma}_u^2$  suit une loi  $\sigma_u^2 \chi_{(n_1+n_2-2)}^2$ , où  $\chi_{(n_1+n_2-2)}^2$  est une variable chi-deux avec  $(n_1 + n_2 - 2)$  degrés de liberté.

Sous l'hypothèse  $H_0$  le test statistique

$$T_c \frac{\hat{\lambda}}{\hat{\sigma}_u \sqrt{1/n_1 + 1/n_2}} \quad (3.12)$$

suit une loi t de Student avec  $(n_1 + n_2 - 2)$  degrés de liberté. L'hypothèse  $H_0: \lambda_T = \lambda_R$  est rejetée en faveur de l'hypothèse alternative  $H_a: \lambda_T \neq \lambda_R$  au seuil de signification  $\alpha$ , si

$$|T_c| > t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right), \quad (3.13)$$

où  $t(\alpha/2, n_1 + n_2 - 2)$  représente la valeur critique supérieure (qui correspond à  $\alpha/2$ ) de la loi t de Student avec  $(n_1 + n_2 - 2)$  degrés de liberté.

En analysant la variance  $\sigma_u^2 = 2(2\sigma_b^2 + \sigma_w^2)$ , on constate qu'elle inclut la variabilité inter-sujet et la variabilité intra-sujet. Dans le cas où la variabilité inter-sujet est relativement grande par rapport à la variabilité intra-sujet, la puissance du test  $T_c$  peut être petite. En pratique, le nombre de sujets d'une étude de bioéquivalence est déterminé en se basant sur la variabilité intra-sujet d'une étude préliminaire. Pour augmenter la puissance du test, Grizzle (1965) suggère de considérer un seuil de signification de 10% et pas de 5% pour tester l'hypothèse nulle (voir Chow et Liu, 2000, chapitre 3).

L'intervalle de confiance  $(1 - \alpha)100\%$  basé sur la statistique  $T_c$  pour  $\lambda$  peut être calculé d'après la formule :

$$\hat{\lambda} \mp t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right) \hat{\sigma}_u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \quad (3.14)$$

Si l'intervalle de confiance contient la valeur 0, l'hypothèse  $H_0$  n'est pas rejetée, et on conclue que l'effet résidu n'est pas significatif.

### 3.3.2 L'effet direct du traitement

En ce qui suit on va tester l'effet direct du traitement.

Soit noté par :

$$d_{ik} = \frac{1}{2}(y_{i2k} - y_{i1k}), \quad (3.15)$$

la différence entre l'observation de la période 2 et l'observation de la période 1 du sujet  $i$  ( $i = 1, 2, \dots, n_k$ ) de la séquence  $k$  ( $k = 1, 2$ ).

Les espérances mathématiques des variables  $d_{ik}$  ( $i = 1, 2, \dots, n_k, k = 1, 2$ ) sont données dans le tableau 3.13 :

**Tableau 3.13** Les espérances mathématiques des variables  $d_{ik}$

Séquence	
1	$E(d_{i1}) = \frac{1}{2}E(y_{i21} - y_{i11}) = \frac{1}{2}[(\pi_2 - \pi_1) + (\tau_T - \tau_R) + \lambda_R]$
2	$E(d_{i2}) = \frac{1}{2}E(y_{i22} - y_{i12}) = \frac{1}{2}[(\pi_2 - \pi_1) + (\tau_R - \tau_T) + \lambda_T]$

La variance des  $d_{ik}$  est :

$$V(d_{ik}) = \sigma_d^2 = \frac{\sigma_w^2}{2}, \quad (i = 1, 2, \dots, n_k, k = 1, 2). \quad (3.16)$$

En vertu des formules (3.3.2) et (3.3.3) on constate que la variance des  $d_{ik}$  comprend seulement la variabilité intra-sujet dans son calcul, d'où l'avantage des plans croisés dans la comparaison des effets des traitements. Cependant, les espérances mathématiques des  $d_{ik}$  comprennent l'effet période et l'effet résidu.

Soit l'effet période et l'effet direct du traitement défini par  $\pi = \pi_2 - \pi_1$  et  $\tau = \tau_T - \tau_R$ , respectivement.

Pour chacune des deux séquences, la moyenne des  $d_{ik}$  ( $i = 1, 2, \dots, n_k$ ) est calculée d'après la formule :

$$\bar{d}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} d_{ik}, \quad k = 1, 2. \quad (3.17)$$

Si on considère le contraste linéaire  $\hat{\tau} = \bar{d}_{.1} - \bar{d}_{.2}$  et on calcule son espérance mathématique, on obtient :

$$\begin{aligned} E(\bar{d}_{.1} - \bar{d}_{.2}) &= E(\bar{d}_{.1}) - E(\bar{d}_{.2}) \\ &= \frac{1}{2} [((\pi_2 - \pi_1) + (\tau_T - \tau_R) + \lambda_R) - ((\pi_2 - \pi_1) + (\tau_R - \tau_T) + \lambda_T)] \\ &= \frac{1}{2} [2(\tau_T - \tau_R) + (\lambda_R - \lambda_T)] = \tau_T - \tau_R + \frac{\lambda_R - \lambda_T}{2} = \tau - \frac{\lambda}{2}, \end{aligned}$$

c'est-à-dire :

$$E(\bar{d}_{.1} - \bar{d}_{.2}) = \tau - \frac{\lambda}{2}, \quad (3.18)$$

où  $\lambda = \lambda_R - \lambda_T$ .

On constate que  $\hat{\tau}$  représente un estimateur non-biaisé du paramètre  $\tau$ , si et seulement si l'effet résidu est nul.

En effet, soit  $\lambda_T = \lambda_R$ , alors

$$\hat{\tau} = \bar{d}_{.1} - \bar{d}_{.2} = \frac{1}{2} [(\bar{y}_{.21} - \bar{y}_{.11}) - (\bar{y}_{.22} - \bar{y}_{.12})] = \bar{y}_T - \bar{y}_R. \quad (3.19)$$

Donc, sous l'hypothèse de l'absence de l'effet résidu,

$$\hat{\tau} = \bar{y}_T - \bar{y}_R$$

est un estimateur non-biaisé du  $\tau$ . Les valeurs

$$\bar{y}_T = \frac{1}{2}(\bar{y}_{.21} + \bar{y}_{.12}) \text{ et } \bar{y}_R = \frac{1}{2}(\bar{y}_{.11} + \bar{y}_{.22})$$

représentent les moyennes obtenues par les moindres carrés (*least square means*) des traitements Teste et Référence, respectivement.

Notons que l'effet direct du traitement,  $\tau$ , est noté également par la différence  $\mu_T - \mu_R$ , où  $\mu_T$  et  $\mu_R$  sont les moyennes des populations du traitement Test et du traitement Référence, respectivement.

Sous les hypothèses du modèle défini pour le plan croisé 2 x 2 (section 3.3),  $\hat{\tau}$  suit une loi normale d'espérance  $\tau$  et de variance

$$V(\hat{\tau}) = \sigma_d^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right). \quad (3.20)$$

Les ensembles  $\{d_{i1}\}$ ,  $i = 1, 2, \dots, n_1$  et  $\{d_{i2}\}$ ,  $i = 1, 2, \dots, n_2$  représentent deux échantillons indépendants tirés de populations normales avec variances égales. Dans le cas où il n'existe pas d'effet résidu, pour tester l'effet direct du traitement on se base sur le test t de Student pour deux échantillons indépendants, de statistique suivante :

$$T_d = \frac{\hat{\tau}}{\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}}, \quad (3.21)$$

où

$$\hat{\sigma}_d^2 = \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^2 \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_{.k})^2. \quad (3.22)$$

Étant donné que  $(n_1 + n_2 - 2)\hat{\sigma}_d^2$  suit une loi  $\sigma_d^2 \chi^2(n_1 + n_2 - 2)$ , la statistique  $T_d$  suit une loi t de Student avec  $n_1 + n_2 - 2$  degrés de liberté. L'intervalle de confiance  $(1 - \alpha)100\%$  pour  $\tau$  peut être calculé d'après la formule :

$$\hat{\tau} \mp t \left( \frac{\alpha}{2}, n_1 + n_2 - 2 \right) \hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}. \quad (3.23)$$

Pour tester la présence de l'effet direct du traitement on définit les hypothèses suivantes :

$$H_0: \tau_R = \tau_T \text{ par rapport } H_a: \tau_R \neq \tau_T. \quad (3.24)$$

L'hypothèse  $H_0$  est rejetée, si

$$|T_d| > t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right). \quad (3.25)$$

Notons que la procédure décrite plus haut teste l'égalité des effets traitement, mais pas la bioéquivalence. L'adaptation des tests de cette section dans le cadre des études de bioéquivalence sera décrite dans la section 3.5.

### 3.3.3 L'effet période

En ce qui suit on teste l'effet période.

Soit définies les différences croisées entre l'observation de la période 2 et l'observation de la période 1 du sujet  $i$  ( $i = 1, 2, \dots, n_1$ ) de la séquence  $k = 1$  et entre l'observation de la période 2 et l'observation de la période 1 du sujet  $i$  ( $i = 1, 2, \dots, n_2$ ) de la séquence  $k = 2$ :

$$w_{ik} = \begin{cases} d_{ik} = \frac{1}{2}(y_{i21} - y_{i11}), & k = 1, \\ -d_{ik} = \frac{1}{2}(y_{i12} - y_{i22}), & k = 2. \end{cases} \quad (3.26)$$

Les espérances mathématiques et la variance des  $w_{ik}$  sont, respectivement :

$$E(w_{ik}) = \begin{cases} \frac{1}{2}[(\pi_2 - \pi_1) + (\tau_T - \tau_R) + \lambda_R] \\ \frac{1}{2}[(\pi_1 - \pi_2) + (\tau_T - \tau_R) - \lambda_T] \end{cases} \quad (3.27)$$

et

$$V(w_{ik}) = \sigma_d^2 = \frac{\sigma_e^2}{2}. \quad (3.28)$$

Soit les moyennes  $\bar{w}_{.1}$  et  $\bar{w}_{.2}$  dans la séquence 1 et 2 définies par :

$$\bar{w}_{.k} = \begin{cases} \bar{d}_{.1}, & k = 1, \\ -\bar{d}_{.2}, & k = 2. \end{cases} \quad (3.29)$$

Le contraste linéaire  $\hat{\pi} = \bar{w}_{.1} - \bar{w}_{.2}$  est un estimateur non-biaisé de l'effet période  $\pi$ . C'est-à-dire

$$\begin{aligned} E[\hat{\pi}] &= E[\bar{w}_{.1} - \bar{w}_{.2}] = E\left[\frac{1}{2}[(\bar{y}_{.21} - \bar{y}_{.11}) - (\bar{y}_{.12} - \bar{y}_{.22})]\right] \\ &= \frac{1}{2}[2(\pi_2 - \pi_1) + (\lambda_R + \lambda_T)] = \pi_2 - \pi_1 = \pi. \end{aligned}$$

Sous la contrainte  $\lambda_R + \lambda_T = 0$ ,  $\hat{\pi}$  est un estimateur non-biaisé de  $\pi$  indépendamment du fait que l'effet résidu est significatif ou pas.

L'intervalle de confiance  $(1 - \alpha) \times 100\%$  pour  $\pi$  peut être calculé d'après la formule :

$$\hat{\pi} \pm t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right) \hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}. \quad (3.30)$$

L'hypothèse  $H_0: \pi_1 = \pi_2$  est rejetée contre l'hypothèse alternative  $H_a: \pi_1 \neq \pi_2$ , si

$$|T_w| > t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right),$$

où

$$T_w = \frac{\hat{\pi}}{\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}}. \quad (3.31)$$

### 3.4 L'analyse de variance pour le plan croisé 2 x 2

Dans les sections précédentes, les inférences statistiques pour les effets fixes du modèle linéaire décrit par les effets du tableau 3.9 ont été déduites en se basant sur des tests  $t$  de Student pour deux échantillons indépendants. Ces tests sont adaptés au cas des tests de

bioéquivalence. Cependant, les mêmes analyses peuvent être réalisées avec la méthode de l'analyse de la variance. On présente cette approche pour que la présentation des plans croisés soit complète.

L'analyse de la variance étudie la dispersion des observations en divisant la somme totale des carrés (SS) en sommes des carrés pour les effets fixes et pour les effets aléatoires.

Soit  $\bar{y} \dots$  la moyenne globale de toutes les  $2(n_1+n_2)$  observations obtenues pour un plan croisé standard  $2 \times 2$ .

Alors la somme totale des carrés pour ce plan d'expérience est définie par la formule :

$$\begin{aligned} SS_{Total} &= \sum_{k=1}^2 \sum_{j=1}^2 \sum_{i=1}^{n_k} (y_{ijk} - \bar{y} \dots)^2 = \sum_{k=1}^2 \sum_{j=1}^2 \sum_{i=1}^{n_k} (y_{ijk} - \bar{y}_{i \cdot k} + \bar{y}_{i \cdot k} - \bar{y} \dots)^2 \\ &= \sum_{k=1}^2 \sum_{j=1}^2 \sum_{i=1}^{n_k} (y_{ijk} - \bar{y}_{i \cdot k})^2 + 2 \sum_{k=1}^2 \sum_{i=1}^{n_k} (\bar{y}_{i \cdot k} - \bar{y} \dots)^2 = SS_W + SS_B, \end{aligned}$$

où :

- $\bar{y}_{i \cdot k} = (1/2) \sum_{j=1}^2 y_{ijk}$ ;
- $SS_B$  représente la somme des carrés due aux sujets (inter-sujet);
- $SS_W$  représente la somme des carrés intra-sujet.

Étant donné que le nombre total d'observations est  $2(n_1+n_2)$ , la somme  $SS_{Total}$  a  $2(n_1+n_2) - 1$  degrés de liberté. Le nombre total de sujets dans les deux séquences est de  $n_1+n_2$ , d'où le nombre de degrés de liberté pour  $SS_B$  et  $SS_W$  sont  $n_1+n_2 - 1$  et  $n_1+n_2$ , respectivement.

La somme  $SS_B$  est sous-divisée elle aussi en deux sommes des carrés, l'une pour l'effet résidu et l'autre pour l'effet aléatoire inter-sujet:

$$SS_B = SS_R + SS_{Inter},$$

où

$$SS_R = \frac{2n_1n_2}{n_1+n_2} [(\bar{y}_{.12} + \bar{y}_{.22}) - (\bar{y}_{.11} + \bar{y}_{.21})]^2,$$

$$SS_{inter} = \sum_{k=1}^2 \sum_{i=1}^{n_k} \frac{y_{i.k}^2}{2} - \sum_{k=1}^2 \frac{y_{..k}^2}{2n_k}.$$

La valeur  $\bar{y}_{..k} = 1/n_k \sum_{i=1}^{n_k} (1/2 \sum_{j=1}^2 y_{ijk})$ .

Les sommes  $SS_R$  et  $SS_{inter}$  ont 1 et  $n_1+n_2 - 2$  degrés de liberté, respectivement. Une somme des carrés divisée par son nombre de degrés de liberté représente une **moyenne des sommes des carrés** (MS).

Pour tester l'effet résidu la statistique suivante est utilisée en analyse de variance :

$$F_c = \frac{MS_{carry}}{MS_{inter}}.$$

Sous l'hypothèse  $H_0: \lambda = 0$  (ou  $\lambda_T = \lambda_R$ ), la statistique  $F_c$  suit une loi  $F$  avec 1 et  $n_1 + n_2 - 2$  degrés de liberté.

L'hypothèse nulle sera rejetée au seuil de signification  $\alpha$ , si

$$F_c > F(\alpha, 1, n_1 + n_2 - 2),$$

où  $F_\alpha$  est le  $\alpha$ -quantile de la loi  $F$ .

Il faut noter que la distribution de Fisher,  $F$ , avec 1 et  $\nu$  degrés de liberté est égale au carré de la loi  $t$  de Student avec  $\nu$  degrés de liberté. Alors la statistique  $F_c$  est équivalente à la statistique  $T_c$  de (3.9), car  $F_c = T_c^2$ .

Notons également que pour le plan croisé  $2 \times 2$  l'effet séquence est confondu avec l'effet résidu. Cela explique la présence de la somme des carrés des résidus comme composante de la somme des carrés due aux sujets,  $SS_B$ .

De façon similaire, la somme  $SS_W$  peut être sous-divisée en trois autres sommes des carrés, c'est-à-dire la somme pour l'effet direct du traitement, la somme pour la période et la somme pour les erreurs intra-sujet, donc

$$SS_W = SS_{Trt} + SS_{Per} + SS_{Intra},$$

où

$$SS_{Trt} = \frac{2n_1n_2}{n_1 + n_2} \{1/2[(\bar{y}_{\cdot 21} - \bar{y}_{\cdot 11}) - (\bar{y}_{\cdot 22} - \bar{y}_{\cdot 12})]\}^2,$$

$$SS_{Per} = \frac{2n_1n_2}{n_1 + n_2} \{1/2[(\bar{y}_{\cdot 21} - \bar{y}_{\cdot 11}) - (\bar{y}_{\cdot 12} - \bar{y}_{\cdot 22})]\}^2,$$

$$SS_{Intra} = \sum_{k=1}^2 \sum_{j=1}^2 \sum_{i=1}^{n_k} y_{ijk}^2 - \sum_{k=1}^2 \sum_{j=1}^2 \frac{y_{\cdot jk}^2}{n_k} + \sum_{k=1}^2 \frac{y_{\cdot \cdot k}^2}{2n_k}.$$

Un degré de liberté est associé à chacune des sommes  $SS_{Trt}$  et  $SS_{Per}$  et  $n_1 + n_2 - 2$  degrés de liberté sont associés à  $SS_{Intra}$ .

Notons que

$$MS_{Intra} = 2\hat{\sigma}_d^2,$$

où  $\hat{\sigma}_d^2$  est de la même formule que (3.11), c'est-à-dire

$$\hat{\sigma}_d^2 = \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^2 \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_{\cdot k})^2.$$

Si l'effet résidu est non significatif,  $\lambda_T = \lambda_R$ , l'hypothèse nulle d'égalité des effets des traitements Test et Référence,  $H_0: \tau_R = \tau_T$ , est testé en utilisant la statistique :

$$F_d = \frac{MS_{Trt}}{MS_{Intra}}.$$

La statistique  $F_d$  suit une loi  $F$  avec 1 et  $n_1 + n_2 - 2$  degrés de liberté. L'hypothèse nulle est rejetée au seuil de signification  $\alpha$ , si

$$F_d > F(\alpha, 1, n_1 + n_2 - 2).$$

La statistique  $F_d$  est équivalente à la statistique  $T_d$  définie par la formule (3.8), c'est-à-dire

$$T_d = \frac{\hat{\tau}}{\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}},$$

car  $F_d = T_d^2$ .

Pour tester l'hypothèse nulle d'égalité des effets période 1 et 2,  $H_0: \pi_1 = \pi_2$ , on peut considérer la statistique

$$F_p = \frac{MS_{per}}{MS_{intra}}.$$

La statistique  $F_p$  suit une loi  $F$  avec 1 et  $n_1 + n_2 - 2$  degrés de liberté. L'hypothèse nulle est rejetée au seuil de signification  $\alpha$ , si

$$F_p > F(\alpha, 1, n_1 + n_2 - 2).$$

La statistique  $F_p$  est équivalente à la statistique  $T_w$  définie par la formule (3.12), c'est-à-dire

$$T_w = \frac{\hat{\pi}}{\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}},$$

car  $F_p = T_w^2$ .

Le tableau 3.14 présenté à la fin de cette section résume l'analyse de la variance pour le plan croisé 2 x 2.

Notons qu'à partir des valeurs  $MS_{inter}$  et  $MS_{intra}$  la présence de la variabilité inter-sujet peut être évaluée en testant l'hypothèse nulle

$$H_0: \sigma_b^2 = 0, \text{ contre } H_a: \sigma_b^2 > 0,$$

où la statistique à utiliser est :

$$F_v = \frac{MS_{Inter}}{MS_{Intra}}$$

La statistique  $F_v$  suit une loi  $F$  avec  $n_1 + n_2 - 2$  et  $n_1 + n_2 - 2$  degrés de liberté sous l'hypothèse nulle. L'hypothèse  $H_0$  est rejetée au seuil de signification  $\alpha$ , si

$$F_v > F(\alpha, n_1 + n_2 - 2, n_1 + n_2 - 2).$$

En conclusion, les plans d'expériences croisés, et surtout le plan croisé  $2 \times 2$ , sont très souvent appliqués dans le cadre des études de bioéquivalence. Leur premier avantage par rapport à d'autres types de plans, est d'être capable d'estimer la variabilité intra-sujet séparément de celle inter-sujet. Cependant, dans le cas où le traitement est potentiellement toxique, ou qu'il a une vie très longue, c'est-à-dire qu'une période de temps relativement longue est nécessaire pour que le traitement soit complètement éliminé de l'organisme, il est préférable d'utiliser le plan d'expérience parallèle. Même si un effet résidu du traitement est encore présent après la fin de la période pendant laquelle ce dernier avait été administré, cela ne cause pas de problème à l'analyse de données. Le plan parallèle ne contient qu'une période. De plus, l'étude aura fini plus vite, ce qui diminue le risque d'abandon des sujets durant l'étude.

Tableau 3.14 Analyse de la Variance

Source de variation	Degrés de libertés (ddl)	SS	$MS = SS/ddl$	$E(MS)$	F
Inter-sujets					
Résidus	1	$SS_R$	$SS_R$	$\frac{2n_1n_2}{n_1+n_2}(\lambda_T - \lambda_R)^2 + 2\sigma_b^2 + \sigma_w^2$	$F_c = MS_R/MS_{Inter}$
Erreur Inter	$n_1 + n_2 - 2$	$SS_{Inter}$	$\frac{SS_{Inter}}{n_1 + n_2 - 2}$		$2\sigma_b^2 + \sigma_w^2$
Intra-sujets					
Traitement	1	$SS_{Trt}$	$SS_{Trt}$	$2n_1n_2/(n_1 + n_2) \times [(\tau_T - \tau_R) + (\lambda_T - \lambda_R)/2]^2 + \sigma_w^2$	$F_d^* = MS_{Trt}/MS_{Intra}$
Période	1	$SS_{per}$	$SS_{per}$		$2n_1n_2/(n_1 + n_2) \times (\pi_2 - \pi_1)^2 + \sigma_w^2$
Erreur Intra	$n_1 + n_2 - 2$	$SS_{Intra}$	$\frac{SS_{Intra}}{n_1 + n_2 - 2}$	$\sigma_w^2$	
Totale	$2(n_1+n_2) - 1$	$SS_{Total}$			

### 3.5 Tests de bioéquivalence

Tel que vu dans le chapitre 1, la bioéquivalence d'un traitement Test et un traitement Référence est testée en fonction du taux et de l'exposition totale de la biodisponibilité. La **biodisponibilité** d'un traitement représente le terme utilisé pour indiquer la quantité totale du traitement (médicament) et la vitesse avec laquelle cette quantité atteint la circulation sanguine. Tel que vu au chapitre 1, le paramètre pharmacocinétique  $AUC$ , qui représente l'aire sous la courbe des concentrations du traitement dans le sang ou le plasma, décrit le mieux l'exposition totale de la biodisponibilité. La concentration maximale du traitement absorbée dans le sang ou le plasma,  $C_{max}$ , et le temps où cette valeur est atteinte,  $t_{max}$ , caractérisent le taux de la biodisponibilité. Donc, une étude de bioéquivalence consiste à comparer le traitement Test par rapport au traitement Référence et à tester leur équivalence par rapport aux paramètres pharmacocinétiques  $AUC$ ,  $C_{max}$  et souvent  $t_{max}$  également.

D'après les règlements des agences qui contrôlent la mise sur le marché des nouveaux médicaments, il suffit de comparer les moyennes des valeurs observées des paramètres pharmacocinétiques pour tester l'équivalence des traitements. Tel que mentionné, ce type d'étude s'appelle **bioéquivalence moyenne**.

Cependant, une différence statistiquement significative entre les mesures de biodisponibilité des deux traitements n'implique pas nécessairement une différence significative de point de vue clinique. Par exemple, en terme de  $AUC$ , une valeur de 80% de la biodisponibilité du traitement Test est obtenue par rapport à la biodisponibilité du traitement Référence. Donc, il existe une différence de 20% entre les  $AUC$  des deux traitements qui peut être statistiquement significative. Cette différence n'est pas nécessairement cliniquement importante en termes d'équivalence thérapeutique. Juste pour se rappeler, deux médicaments (traitements) sont considérés équivalents de point de vue thérapeutique, s'ils contiennent la même quantité du même ingrédient actif du traitement et, de plus, ils donnent le même effet thérapeutique. Malgré la différence de 20% entre les

valeurs de  $AUC$  du traitement Test et du traitement Référence ces deux traitements peuvent avoir le même effet thérapeutique. Donc, dans ce cas, ils doivent être considérés équivalents de point de vue thérapeutique.

Comme déjà mentionné dans le chapitre 2, les hypothèses de bioéquivalence des deux traitements, Test et Référence, se définissent par:

$$H_0: \mu_T - \mu_R < \theta_0 - \varepsilon_1 \quad \text{ou} \quad \mu_T - \mu_R > \theta_0 + \varepsilon_2 \quad (3.32)$$

$$\text{contre } H_a: \theta_0 - \varepsilon_1 \leq \mu_T - \mu_R \leq \theta_0 + \varepsilon_2,$$

où  $\mu_T$  et  $\mu_R$  sont les moyennes des traitements Test et Référence, respectivement, et  $\theta_0 - \varepsilon_1$ ,  $\theta_0 + \varepsilon_2$  sont des limites cliniquement pertinentes définies a priori. Rappelons-nous que la différence  $\mu_T - \mu_R$  n'est que l'effet direct du traitement,  $\tau$  (voir section 3.3.2) et que par le test décrit à la section 3.3.2 on ne teste que l'égalité des effets traitements, mais pas leur équivalence.

En pratique, ce qu'on appelle **la règle  $\pm 20$**  proposée par l'agence qui contrôle la mise sur le marché des nouveaux médicaments aux États Unis, FDA (Food and Drug Administration), est appliquée, généralement, pour les études de bioéquivalence. D'après cette règle, la bioéquivalence est conclue, si la biodisponibilité moyenne du traitement Test n'excède pas les limites de  $\pm 20\%$  de la biodisponibilité du traitement Référence. En d'autres termes, le rapport des moyennes  $\mu_T/\mu_R$  ne doit pas dépasser les limites  $[0.80; 1.20]$ . D'après la règle  $\pm 20$ ,

$$-0.20\mu_R \leq \mu_T - \mu_R \leq 0.20\mu_R,$$

$$-0.20 \leq \frac{\mu_T - \mu_R}{\mu_R} \leq 0.20,$$

$$-0.20 \leq \frac{\mu_T}{\mu_R} - 1 \leq 0.20,$$

$$0.80 \leq \frac{\mu_T}{\mu_R} \leq 1.20,$$

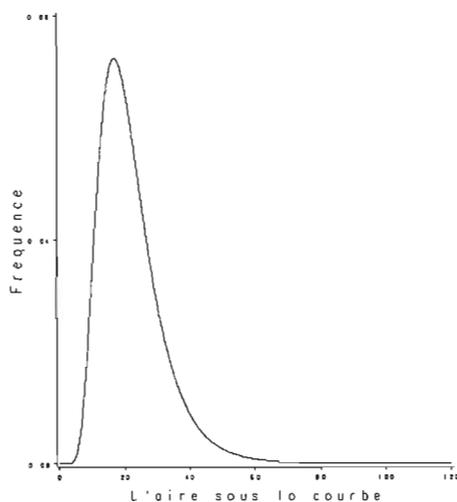
où  $\mu_T$  est la moyenne du traitement Test et  $\mu_R$  est la moyenne du traitement Référence. Donc,  $\theta_0 - \varepsilon_1 = -0.20\mu_R$  et  $\theta_0 + \varepsilon_2 = 0.20\mu_R$ . En termes du rapport de la moyenne du traitement Test et de la moyenne du traitement Référence le test de bioéquivalence est décrit par les hypothèses :

$$H_0: \frac{\mu_T}{\mu_R} < \delta_L \quad \text{ou} \quad \frac{\mu_T}{\mu_R} > \delta_U \quad (3.33)$$

$$\text{contre } H_a: \delta_L \leq \frac{\mu_T}{\mu_R} \leq \delta_U,$$

où  $\delta_L = 0.80$  et  $\delta_U = 1.20$ .

Notons que, d'après leur définition, les paramètres  $AUC$  et  $C_{max}$  ne prennent que des valeurs positives ( $\geq 0$ ) et qu'ils ne suivent pas des lois normales, mais plutôt des lois log normales comme dans la figure 3.1.



**Figure 3.1** Loi log normale

Dans la pratique courante, les valeurs de ces paramètres sont souvent transformées en logarithme naturel dans le but d'obtenir des valeurs qui suivent des lois normales. Dans le

tableau 3.15 sont donnés certains paramètres importants de la loi normale et de la loi log-normale, respectivement.

**Tableau 3.15** Paramètres importants de la loi normale et de la loi log-normale

	Loi normale	Loi log normale
<b>Moyenne</b>	$\mu$	$e^{\mu+\sigma^2/2}$
<b>Variance</b>	$\sigma^2$	$e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$
<b>Coefficient de variation (CV)</b>	$\sigma/\mu$	$\sqrt{e^{\sigma^2} - 1}$

Le paramètre  $t_{max}$  n'est pas transformé, habituellement.

Soit  $\ln X_1, \ln X_2, \dots, \ln X_{n_1}$  et  $\ln Y_1, \ln Y_2, \dots, \ln Y_{n_2}$  les valeurs transformées d'un des paramètres pharmacocinétiques ( $AUC$  ou  $C_{max}$ , par exemple). Alors, les moyennes échantillonales des valeurs logarithmiques sont :

$$\bar{X}_{\ln} = \sum_{i=1}^{n_1} \frac{\ln(X_i)}{n_1} \text{ et } \bar{Y}_{\ln} = \sum_{j=1}^{n_2} \frac{\ln(Y_j)}{n_2},$$

respectivement. Pour représenter les résultats à l'échelle non-transformée on transforme ces moyennes en exponentielle, c'est-à-dire, on considère

$$e^{\bar{X}_{\ln}} \text{ et } e^{\bar{Y}_{\ln}}.$$

Donc comme résultat on obtient des moyennes géométriques.

Si les hypothèses définies par (3.33) sont formulées en termes de logarithme naturel, on obtient le test d'hypothèses pour les valeurs log-transformées (voir 3.34). Notons que ces hypothèses sont équivalentes à celles du (3.32) définies pour des observations non transformées. Donc, on considère

$$H_0: \ln(\mu_T) - \ln(\mu_R) < \ln \delta_L \quad \text{ou} \quad \ln(\mu_T) - \ln(\mu_R) > \ln \delta_U \quad (3.34)$$

$$\text{contre } H_a: \ln \delta_L \leq \ln(\mu_T) - \ln(\mu_R) \leq \ln \delta_U,$$

étant donné que

$$\ln\left(\frac{\mu_T}{\mu_R}\right) = \ln(\mu_T) - \ln(\mu_R).$$

Cependant, pour les valeurs transformées les limites de bioéquivalence sont fixées à  $[\ln 0.80, \ln 1.25]$  par l'agence responsable de la mise des médicaments sur le marché aux États-Unis, FDA. La raison de cela est,

$$\ln(0.80) \leq \ln(\mu_T) - \ln(\mu_R) \leq \ln(1.25),$$

d'où

$$-0.22 \leq \ln(\mu_T) - \ln(\mu_R) \leq 0.22.$$

Tandis que, si on avait pris

$$\ln(0.80) \leq \ln(\mu_T) - \ln(\mu_R) \leq \ln(1.20),$$

cela aurait donné

$$-0.22 \leq \ln(\mu_T) - \ln(\mu_R) \leq 0.18$$

et on préfère avoir un changement de signes pour les bornes de la différence.

$$\ln(\mu_T) - \ln(\mu_R) \text{ et } \ln(\mu_R) - \ln(\mu_T).$$

Rappelons-nous que les valeurs des paramètres pharmacocinétiques sont souvent transformées en logarithme naturel dans le but d'obtenir des valeurs qui suivent des lois

normales. Soit les valeurs du paramètre  $AUC$ , par exemple, transformées en logarithme naturel. Notons par

$$\bar{X}_{ln} = \sum_{i=1}^{n_1} \frac{\ln(x_i)}{n_1} = \bar{U} \sim \mathcal{N}(\mu_U, \sigma^2),$$

$$\bar{Y}_{ln} = \sum_{j=1}^{n_2} \frac{\ln(Y_j)}{n_2} = \bar{V} \sim \mathcal{N}(\mu_V, \sigma^2),$$

les moyennes échantillonales des valeurs logarithmiques.

Leurs espérances mathématiques sont égales à :

$$E[\bar{X}_{ln}] = E[\bar{U}] = \mu_U \text{ et } E[\bar{Y}_{ln}] = E[\bar{V}] = \mu_V.$$

Si on transforme ces valeurs en exponentielle, on obtient :

$$e^{E[\bar{U}]} = e^{\mu_U} \text{ et } e^{E[\bar{V}]} = e^{\mu_V}. \quad (3.35)$$

Toutefois, en pratique on ramène les valeurs des moyennes  $\bar{X}_{ln}$ ,  $\bar{Y}_{ln}$  à l'échelle non transformée en faisant  $e^{\bar{X}_{ln}} = e^{\bar{U}}$  et  $e^{\bar{Y}_{ln}} = e^{\bar{V}}$ .

Notons leurs espérances mathématiques par :

$$\mu_T = E[e^{\bar{X}_{ln}}] = E[e^{\bar{U}}] \text{ et } \mu_R = E[e^{\bar{Y}_{ln}}] = E[e^{\bar{V}}]. \quad (3.36)$$

Remarquons qu'on n'a pas d'égalité entre les valeurs des formules (3.35) et les valeurs des formules (3.36), c'est-à-dire que :

$$E[e^{\bar{U}}] \neq e^{E[\bar{U}]} \text{ et } E[e^{\bar{V}}] \neq e^{E[\bar{V}]}.$$

Donc, on constate que :

$$\mu_T \neq e^{\mu_U} \text{ et } \mu_R \neq e^{\mu_V}. \quad (3.37)$$

Par une transformation en logarithme naturel on obtient les inégalités :

$$\ln(\mu_T) \neq \ln(e^{\mu_U}) = \mu_U \text{ et } \ln(\mu_R) \neq \ln(e^{\mu_V}) = \mu_V.$$

Cependant, on veut savoir si l'égalité suivante n'est pas erronée, c'est-à-dire si on a

$$\ln(\mu_T) - \ln(\mu_R) = \mu_U - \mu_V. \quad (3.38)$$

Commençons par calculer

$$\ln(\mu_T) - \ln(\mu_R) = \ln\left(\frac{\mu_T}{\mu_R}\right) = \ln\left(\frac{E[e^{\bar{U}}]}{E[e^{\bar{V}}]}\right).$$

Comme

$$\bar{U} \sim \mathcal{N}(\mu_U, \sigma^2) \text{ et } \bar{V} \sim \mathcal{N}(\mu_V, \sigma^2),$$

alors

$$\bar{U} - \mu_U \sim \mathcal{N}(0, \sigma^2) \text{ et } \bar{V} - \mu_V \sim \mathcal{N}(0, \sigma^2),$$

respectivement. De plus, leurs espérances mathématiques sont égales, car  $\bar{U} - \mu_U$  et  $\bar{V} - \mu_V$  suivent des lois normales de même moyenne égale à zéro et de même variance. Donc,

$$\frac{E[e^{\bar{U}}]}{E[e^{\bar{V}}]} = \frac{E[e^{\bar{U} - \mu_U + \mu_U}]}{E[e^{\bar{V} - \mu_V + \mu_V}]} = \frac{E[e^{\bar{U} - \mu_U}]e^{\mu_U}}{E[e^{\bar{V} - \mu_V}]e^{\mu_V}} = \frac{e^{\mu_U}}{e^{\mu_V}} = e^{\mu_U - \mu_V}.$$

On obtient,

$$\frac{E[e^{\bar{U}}]}{E[e^{\bar{V}}]} = e^{\mu_U - \mu_V} = e^{E[\bar{U}] - E[\bar{V}]},$$

c'est-à-dire

$$\frac{\mu_T}{\mu_R} = e^{\mu_U - \mu_V}$$

et donc,

$$\ln\left(\frac{\mu_T}{\mu_R}\right) = \mu_U - \mu_V,$$

d'où

$$\ln(\mu_T) - \ln(\mu_R) = \mu_U - \mu_V.$$

Donc, malgré que  $\ln(\mu_T) \neq \mu_U$  et  $\ln(\mu_R) \neq \mu_V$  les différences de ces moyennes sont égales et l'égalité (3.38) est satisfaite. De plus, même, si on a supposé que les échantillons sont de lois normales, ces raisonnements se généralisent à deux échantillons de même loi et de même variance.

D'après la procédure proposée par Schuirmann (1981, 1987), le test de bioéquivalence défini par les hypothèses (3.32) peut s'effectuer en considérant deux tests unilatéraux d'équivalence.

On obtient donc,

$$H_{01}: \mu_T - \mu_R \leq \theta_0 - \varepsilon_1 \text{ contre } H_{a1}: \mu_T - \mu_R > \theta_0 - \varepsilon_1 \quad (3.39)$$

et

$$H_{02}: \mu_T - \mu_R \geq \theta_0 + \varepsilon_2 \text{ contre } H_{a2}: \mu_T - \mu_R < \theta_0 + \varepsilon_2 \quad (3.40)$$

Par les hypothèses (3.39) on vérifie si la biodisponibilité du traitement Test n'est pas inférieure à la biodisponibilité du traitement Référence (hypothèse alternative). Par les hypothèses (3.40), par contre, on teste si la biodisponibilité du traitement Test n'est pas supérieure à celle du traitement Référence (hypothèse alternative).

Notons qu'on a émis les hypothèses pour le cas où les données sont non-transformées. Toutefois, cette procédure est appliquée de la même façon au cas où on transforme les données, en remplaçant  $\mu_T$ ,  $\mu_R$  par leur logarithme (voir 3.34).

D'après la procédure des tests unilatéraux d'équivalence, on constate l'équivalence des moyennes  $\mu_T$  et  $\mu_R$  au seuil de signification de  $\alpha\%$ , si et seulement si les hypothèses nulles  $H_{01}$  et  $H_{02}$  sont rejetées simultanément au seuil de signification de  $\alpha\%$ .

Notons que la procédure composée des deux tests unilatéraux pour tester la bioéquivalence entre deux moyennes au seuil de signification de  $\alpha\%$  se réduit à un intervalle de confiance à  $(1-2\alpha)100\%$ .

En effet, sous l'hypothèse de normalité des données, les deux ensembles d'hypothèses unilatérales d'équivalence sont testés par des tests t de Student unilatéraux. Tel que vu dans le chapitre 2, section 2.4, le principe de raisonnement pour un test unilatéral d'équivalence est le même que pour un test unilatéral ordinaire.

De plus, les moyennes  $\mu_T$  et  $\mu_R$  sont équivalentes, si les statistiques  $T_L$  et  $T_U$  satisfont simultanément les inégalités suivantes :

$$T_L = \frac{(\bar{Y}_T - \bar{Y}_R) - (\theta_0 - \varepsilon_1)}{\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}} > t(\alpha, n_1 + n_2 - 2) \quad (3.41)$$

et

$$T_U = \frac{(\bar{Y}_T - \bar{Y}_R) - (\theta_0 + \varepsilon_2)}{\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}} < -t(\alpha, n_1 + n_2 - 2).$$

Tel que définit à la section 3.3.2, les valeurs  $\bar{Y}_T$  et  $\bar{Y}_R$  sont les moyennes des moindres carrés du traitement Test et du traitement Référence, respectivement. On a aussi l'estimateur  $\hat{\sigma}_d$  de la variance  $\sigma_d$  des différences  $d_{ik}$  entre l'observation de la période 2 et l'observation de la période 1 du sujet  $i$  ( $i = 1, 2, \dots, n_k$ ) de la séquence  $k$  ( $k = 1, 2$ ) pour un plan d'expérience croisé  $2 \times 2$ .

Les deux dernières statistiques (3.41) peuvent être également écrites sous la forme :

$$\theta_0 - \varepsilon_1 \leq \bar{Y}_T - \bar{Y}_R - t(\alpha, n_1 + n_2 - 2) \left[ \hat{\sigma}_d \sqrt{1/n_1 + 1/n_2} \right] \quad (3.42)$$

et

$$\theta_0 + \varepsilon_2 \geq \bar{Y}_T - \bar{Y}_R - t(\alpha, n_1 + n_2 - 2) \left[ \hat{\sigma}_d \sqrt{1/n_1 + 1/n_2} \right].$$

Elles sont aussi équivalentes à :

$$\theta_0 - \varepsilon_1 \leq (\bar{Y}_T - \bar{Y}_R) \pm t(\alpha, n_1 + n_2 - 2) [\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}] \leq \theta_0 + \varepsilon_2. \quad (3.43)$$

La dernière expression (3.43) n'est que l'intervalle classique de confiance à  $(1 - 2\alpha)100\%$ .

Donc, on conclut que la procédure composée des deux tests unilatéraux pour tester la bioéquivalence entre deux moyennes au seuil de signification de  $\alpha\%$  se réduit à un intervalle de confiance à  $(1 - 2\alpha)100\%$ .

Notons que l'intervalle de confiance est construit à 90%, par exemple, tandis que le seuil de signification reste à  $\alpha = 5\%$  et pas à  $\alpha = 10\%$ . En effet, tel que vu dans le chapitre 2, section 2.5, sous  $H_0$  on a soit  $\mu_T - \mu_R < \theta_0 - \varepsilon_1$ , soit  $\mu_T - \mu_R > \theta_0 + \varepsilon_2$  et la différence des moyennes ne peut pas être dans les deux intervalles en même temps. Donc, la probabilité de commettre une erreur de type I reste de 5% d'un côté de l'intervalle  $[\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2]$  et 0% pour l'autre côté de l'intervalle  $[\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2]$ , ce qui donne en total 5% et pas 10%.

Si les valeurs des paramètres pharmacocinétiques sont transformées en logarithme naturel, l'intervalle de confiance est facilement transformé dans un intervalle de confiance pour le rapport  $(\mu_T/\mu_R)$  :

$$e^{(\bar{U} - \bar{V} \pm t(\alpha, n_1 + n_2 - 2) [\hat{\sigma}_d \sqrt{1/n_1 + 1/n_2}])}. \quad (3.44)$$

Dans notre étude de simulation on va appliquer le principe du Schuirman (1981, 1987) des deux tests unilatéraux pour répondre à une question de test bilatéral.

## CHAPITRE IV

### ÉTUDE DE SIMULATION

#### 4.1 Description de l'expérience et de ses paramètres

Généralement, les études de bioéquivalence sont basées sur des échantillons de petite taille dont les sujets sont attribués aléatoirement aux traitements. En pratique, on utilise le test t de Student pour réaliser l'inférence statistique dans le cadre de ces études. Généralement, l'une des hypothèses importantes sous lesquelles ce test paramétrique est appliqué est la normalité de la loi des observations ou d'une transformation de celle-ci. Dans ce chapitre, on applique le test t de Student aux données transformées.

En deuxième, considérons le test de permutations (chapitre 2), un test non paramétrique qui n'exige aucune hypothèse sur la loi des données, à part le fait que les sujets soient attribués aléatoirement aux traitements. En recalculant la statistique pour toutes les permutations possibles de l'échantillon observé, le test de permutations génère lui-même la loi de référence.

En troisième, considérons le test de Wilcoxon pour deux échantillons (chapitre 2) qui est un test non paramétrique libre de loi lui aussi, mais qui est basé sur les rangs des observations et pas sur les valeurs elles-mêmes, contrairement au test de permutations. Parmi les trois tests mentionnés ci-dessus, ce dernier est le test le plus conservateur (qui rejette le moins souvent).

Dans ce chapitre, on simule des données de plusieurs types de loi. On applique le test des permutations sur ces données simulées et ensuite on compare les résultats obtenus avec ce test aux résultats du test t de Student et du test de Wilcoxon pour deux échantillons.

Dans notre cas, on suppose que les observations simulées représentent des valeurs  $\ln(AUC)$  générées selon le modèle de plan croisé 2 x 2 défini dans le chapitre 3.

Les effets fixes et aléatoires du modèle sont initialisés de la façon suivante :

- les effets de la période,  $\pi_1$ ,  $\pi_2$ , et les effets résidus dû au traitement Test et au traitement Référence,  $\lambda_T$ ,  $\lambda_R$ , respectivement, sont considérés nuls ;
- les effets aléatoires des deux séquences,  $S_{i1}$ ,  $S_{i'2}$ , suivent une loi normale  $\mathcal{N}(0, \sigma_B^2)$ ;
- en ce qui concerne les erreurs expérimentales,  $e_{ijk}$ , on étudie plusieurs situations:

**Situation 1.** les erreurs suivent une loi normale  $\mathcal{N}(0, \sigma_W^2)$ ;

**Situation 2.** les erreurs suivent une loi de mélange composée d'une loi normale  $\mathcal{N}(0, \sigma_W^2)$  et d'une loi double-exponentielle,  $\mathcal{DE}(0, \sigma_W/\sqrt{2})$ , les deux d'espérance zéro et de variance  $\sigma_W^2$ . La fonction de répartition de cette loi est donnée par la formule:

$$F(x) = pF_1(x) + (1 - p)F_2(x),$$

où  $p$  est la proportion de la loi normale caractérisée par la fonction de répartition  $F_1$  et  $(1 - p)$  est la proportion de la loi double-exponentielle caractérisée par la fonction de répartition  $F_2$  ;

**Situation 3.** les erreurs suivent une loi double-exponentielle  $\mathcal{DE}(0, \sigma_W/\sqrt{2})$ , de moyenne zéro et de variance  $\sigma_W^2$ .

Afin de tester la bioéquivalence des traitements au seuil de signification de  $\alpha = 5\%$ , on utilise la procédure des deux tests unilatéraux d'équivalence définis par les hypothèses

$$H_{01}: \ln(\mu_T) - \ln(\mu_R) \leq \ln(0.80) \tag{4.1.}$$

$$\text{contre } H_{a1}: \ln(\mu_T) - \ln(\mu_R) > \ln(0.80)$$

et

$$H_{02}: \ln(\mu_T) - \ln(\mu_R) \geq \ln(1.25) \quad (4.2.)$$

contre  $H_{a2}: \ln(\mu_T) - \ln(\mu_R) < \ln(1.25)$ .

Tel que mentionné, dans ce chapitre on analyse les résultats obtenus en appliquant le test de permutation sur des données simulées d'après le modèle du plan d'expérience croisé 2 x 2, dont les erreurs expérimentales suivent plusieurs types de lois : la loi normale, la loi de mélange (la loi normale contaminée par la loi double-exponentielle) et la loi double-exponentielle. De plus, pour chacun de ces cas, le rapport des moyennes du traitement Test et du traitement Référence,  $\mu_T/\mu_R$ , est considéré égal à 80%, 90%, 100%, 110% et 125%. Rappelons-nous que, pour une étude de bioéquivalence, l'équivalence des deux traitements est définie par l'hypothèse alternative. Pour les observations des paramètres pharmacocinétiques log transformées l'intervalle de bioéquivalence est de [80%,125%] ce qui revient à [80%,120%] pour les données non transformées. De plus, en transformant exponentiellement la différence des moyennes des moindres carrées des traitements obtenues suite à une analyse de bioéquivalence sur des données qui suivent une loi log normale, on obtient un rapport des moyennes géométriques des traitements Test et Référence (noté dans cette section par  $\mu_T/\mu_R$ ).

Donc, dans la section des résultats on étudie les erreurs de type I et les puissances qu'on obtient avec le test de permutations aux bornes de l'intervalle de bioéquivalence (80% et 125%) et pour le rapport des moyennes  $\mu_T/\mu_R$  égale à: 90%, 100%, 110%, respectivement. Ensuite, on compare ces valeurs à ceux obtenus par un test paramétrique et un test non paramétrique, c'est-à-dire par le test t de Student et le test de Wilcoxon pour deux échantillons.

En ce qui suit, on décrit la procédure utilisée pour simuler des variables aléatoires. Pour réaliser la partie pratique, on a utilisé le progiciel SAS, version 9.1.3.

#### 4.2 Simulation des erreurs expérimentales

Le progiciel SAS contient déjà des fonctions pour simuler des variables aléatoires pour certaines lois discrètes ou continues comme : la loi binomiale, uniforme, normale, etc. Pour générer des données de loi normale  $\mathcal{N}(0,1)$ , on utilise la fonction SAS,  $rannor(seed)$ , où  $seed$  est le germe. Dans le cas où les données suivent la loi  $\mathcal{N}(\mu, \sigma^2)$ , on utilise la formule :

$$x = \mu + \sigma * rannor(seed).$$

Cependant, il n'existe pas de telles fonctions prédéfinies pour simuler une variable aléatoire de loi double-exponentielle, par exemple, et, particulièrement, pour une loi de mélange. En ce qui suit on décrit la démarche utilisée pour créer des outils afin de simuler des variables aléatoires de loi double-exponentielle  $\mathcal{DE}(0,1)$  définie par la fonction de répartition suivante :

$$F(x) = \begin{cases} \frac{1}{2}e^x, & x < 0 \\ 1 - \frac{1}{2}e^{-x}, & x \geq 0 \end{cases}, \quad (4.3)$$

dont la fonction de densité est :

$$f(x) = \frac{1}{2}e^{-|x|} = \begin{cases} \frac{1}{2}e^{-x}, & x \geq 0 \\ \frac{1}{2}e^x, & x < 0 \end{cases}.$$

Par la suite, on parle comment on simule des variables de la loi de mélange définie par la fonction de répartition donnée par la formule:

$$F(x) = pF_1(x) + (1 - p)F_2(x), \quad (4.4)$$

où  $F_1$  est la fonction de répartition de loi normale  $\mathcal{N}(0, \sigma_w^2)$  et  $F_2$  est la fonction de répartition de loi double-exponentielle  $\mathcal{DE}(0, \sigma_w/\sqrt{2})$ , les deux d'espérance zéro et de variance  $\sigma_w^2$ .

Notons que la variance d'une loi double-exponentielle est égale à  $2\sigma^2$  et dans notre cas  $\sigma = \sigma_W/\sqrt{2}$ , donc

$$2\sigma^2 = 2(\sigma_W/\sqrt{2})^2 = \sigma_W^2.$$

Soit la variable aléatoire  $X$  qui suit une loi double-exponentielle de fonction de répartition définie dans (4.3.). Par définition, la probabilité

$$P(|X| \leq x) = P(-x \leq X \leq x) = F(x) - F(-x) = 1 - \frac{1}{2}e^{-x} - \frac{1}{2}e^{-x} = 1 - e^{-x}.$$

Étant donné que la loi double-exponentielle  $\mathcal{DE}(0,1)$  est symétrique par rapport à zéro, alors les probabilités suivantes sont égales :

$$P(X > 0) = P(X < 0) = 1/2.$$

Dans ces conditions, on a :

$$\begin{aligned} P(|X| \leq x, X > 0) &= P(-x \leq X \leq x, X > 0) = F(x) - F(0) = 1 - \frac{1}{2}e^{-x} - \frac{1}{2} = \frac{1}{2}(1 - e^{-x}) \\ &= P(|X| \leq x)P(X > 0). \end{aligned}$$

Donc pour générer une variable aléatoire de loi double exponentielle, il suffit de générer une variable aléatoire de fonction de répartition :

$$F(x) = 1 - e^{-x} \tag{4.5.}$$

et de l'affecter par un signe «+» ou «-» simulé par la loi Bernoulli,  $Ber(1/2)$ , de probabilité de succès  $p = 1/2$  de la façon suivante :

$$2B - 1 = \begin{cases} -1, & \text{si } B = 0 \\ 1, & \text{si } B = 1 \end{cases}$$

où  $B \sim Ber(1/2)$  est une variable aléatoire Bernoulli.

Tel que mentionné, il existe déjà une fonction en SAS pour simuler des variables aléatoires binomiales,  $\text{ranbin}(\text{seed}, n, p)$ . Donc on peut générer une variable Bernoulli à partir de la loi binomiale  $\text{Bin}(1, p)$  avec la fonction SAS  $\text{ranbin}(\text{seed}, 1, p)$ .

Pour simuler des variables aléatoires,  $X$ , de fonction de répartition (4.5.), on peut créer notre propre fonction basée sur le théorème classique suivant :

*Soit  $U$  une variable aléatoire uniforme  $(0,1)$ . Pour toute fonction de répartition continue  $F$  la variable aléatoire  $X$  défini par  $X = F^{-1}(U)$  est de loi dont la fonction de répartition est  $F$ . Ainsi,  $F^{-1}(u)$  est égale à la valeur  $x$  et  $F(x) = u$ .*

On pose, donc,

$$u = F(x) = 1 - e^{-x}, \quad (4.6.)$$

où  $u$  est une variable aléatoire de loi uniforme  $\mathcal{U}(0,1)$ . Alors,

$$1 - u = e^{-x} \text{ et}$$

$$x = F^{-1}(u) = -\log(1 - u).$$

On conclut que pour générer des variables aléatoires de loi double-exponentielle  $\mathcal{DE}(0,1)$  on suit les étapes suivantes.

**Étape 1.** on génère des variables aléatoires Bernoulli,  $B_i \sim \text{Ber}\left(\frac{1}{2}\right)$  ;

**Étape 2.** on génère des variables aléatoires uniformes,  $U_i \sim \mathcal{U}(0,1)$  ;

**Étape 3.** la variable aléatoire  $X_i$  de loi double-exponentielle  $\mathcal{DE}(0,1)$  est obtenue par :

$$X_i = (2B_i - 1)(-\log(1 - U_i)).$$

Notons qu'on génère une variable uniforme  $\mathcal{U}(0,1)$  avec la fonction SAS  $\text{ranuni}(\text{seed})$ . Pour générer une variable aléatoire de loi double-exponentielle  $\mathcal{DE}(0, \sigma_w/\sqrt{2})$ , on remplace l'Étape 3 par l'Étape 4 suivante :

**Étape 4.** la variable aléatoire  $X_i$  de loi double-exponentielle  $\mathcal{DE}(0, \sigma_w/\sqrt{2})$  est obtenue par :

$$X_i = (2B_i - 1)(-\sigma_w/\sqrt{2})\log(1 - U_i).$$

En effet, si  $x$  dans la formule (4.6.) est remplacé par :

$$y = \frac{x - \mu}{b},$$

alors on obtient la fonction :

$$y = \frac{x - \mu}{\sigma} = F^{-1}(u) = -\log(1 - u),$$

d'où on constate que :

$$x = \mu - \sigma \log(1 - u).$$

Dans notre cas  $\mu = 0$  et  $\sigma = \sigma_w/\sqrt{2}$ , donc la valeur du  $x$  est calculée par :

$$x = -\frac{\sigma_w}{\sqrt{2}}\log(1 - u).$$

Soit  $X$  une variable aléatoire de loi de mélange, dont la fonction de répartition est définie par :

$$F(x) = pF_1(x) + (1 - p)F_2(x).$$

En termes de probabilité, cette fonction peut être écrite sous la forme :

$$P(X \leq x) = P(X \leq x|B = 1)P(B = 1) + P(X \leq x|B = 0)P(B = 0),$$

où  $B \sim \text{Ber}(p)$  est une variable Bernoulli et  $P(B = 1) = p$  et  $P(B = 0) = 1 - p$ . Pour générer une variable aléatoire de loi de mélange, on suit les étapes suivantes :

**Étape 1.** on génère des variables aléatoires Bernoulli,  $B_i \sim \text{Ber}(p)$  ;

**Étape 2.** si  $B_i = 1$ , on génère une variable aléatoire  $X_i$  de loi normale  $\mathcal{N}(0, \sigma_w^2)$  ; si

$B_i = 0$ , on génère une variable aléatoire  $X_i$  de loi double-exponentielle

$\mathcal{DE}(0, \sigma_w/\sqrt{2})$ .

Pour ce qui est du code utilisé pour simuler des données du modèle du plan croisé 2 x 2 avec des effets aléatoires  $S_{ik}$ , ( $i = 1, 2, \dots, n$ ;  $k = 1, 2$ ) et des erreurs expérimentales  $e_{ijk}$ , pour une des lois énumérées dans la section 4.1, il est présenté dans l'appendice B et se base sur l'équation (défini dans le chapitre 3, section 3.3) qui décrit le plan d'expérience croisé 2 x 2.

#### 4.3 Tests effectués

En ce qui suit, on décrit les étapes utilisées pour réaliser les tests de bioéquivalence.

**Étape 1.** D'après le modèle du plan croisé 2 x 2 on simule une base de  $2n$  données avec  $n$  données dans chaque séquence.

Notons que le nombre de sujets dans la séquence 1 peut être différent du nombre dans la séquence 2. Cependant, il est préférable que ce nombre de sujets dans les séquences soit équilibré.

**Étape 2.** On réduit le plan croisé 2 x 2 à deux échantillons indépendants qui représentent les données des séquences :  $TR$  (séquence 1) et  $RT$  (séquence 2), en appliquant la procédure décrite dans le chapitre 3, section 3.3.2. Cette procédure consiste à soustraire l'observation de la période 2 de l'observation de la période 1 du sujet  $i$  ( $i = 1, 2, \dots, n$ ) de la séquence  $k$  ( $k = 1, 2$ ), c'est-à-dire :

$$d_{ik} = \frac{y_{i1k} - y_{i2k}}{2},$$

où, tel que mentionné, on suppose que  $y_{ijk}$  sont des valeurs observées de la variable  $\ln(AUC)$ .

Les échantillons obtenus sont indépendants et ils se l'un de l'autre par la différence du logarithme naturel de la moyenne du traitement Test et du logarithme naturel de la moyenne du traitement Référence,  $\ln(\mu_T/\mu_R) = \ln(\mu_T) - \ln(\mu_R)$ . En effet, tel que vu dans le chapitre 3, sections 3.3 et 3.5, l'espérance mathématique de la différence entre l'observation de la période 1 et l'observation de la période 2 du sujet  $i$  ( $i = 1, 2, \dots, n$ ) de la séquence 1 ou 2 peut

être réduite aux formules données dans le tableau 4.1, car pour notre étude de simulation on considère l'effet période  $(\pi_1, \pi_2)$  et l'effet résidu  $(\lambda_T, \lambda_R)$  nul.

**Tableau 4.1** L'espérance mathématique de la différence entre l'observation de la période 1 et l'observation de la période 2 du sujet  $i$  ( $i = 1, 2, \dots, n$ ) dans la séquence 1 ou 2.

Séquence	
1	$(1/2)E(y_{i11} - y_{i21}) = (1/2)(\ln(\mu_T) - \ln(\mu_R)),$
2	$(1/2)E(y_{i12} - y_{i22}) = (1/2)(\ln(\mu_R) - \ln(\mu_T)).$

Pour calculer l'effet direct du traitement, de la valeur obtenue pour la séquence 1 (voir tableau 4.1) on soustrait la valeur obtenue pour la séquence 2 (voir procédure de la section 3.3). Cela, tel que mentionné, donne la différence suivante :

$$\frac{\ln(\mu_T) - \ln(\mu_R)}{2} - \frac{\ln(\mu_R) - \ln(\mu_T)}{2} = \ln(\mu_T) - \ln(\mu_R) = \ln\left(\frac{\mu_T}{\mu_R}\right).$$

**Étape 3.** Pour le test unilatéral d'équivalence (4.1.) on soustrait la constante  $\ln(80)/2$  des observations de la séquence 1 et on rajoute cette constante aux observations de la séquence 2. Cela s'explique par le raisonnement suivant.

Le test unilatéral d'équivalence (4.1) peut être également écrit sous la forme :

$$H_{01}: \ln(\mu_T) - \ln(\mu_R) - \ln(0.80) \leq 0,$$

$$\text{contre } H_{a1}: \ln(\mu_T) - \ln(\mu_R) - \ln(0.80) > 0.$$

Cette différence définie par les hypothèses du test, peut être décomposée comme suit:

$$\ln(\mu_T) - \ln(\mu_R) - \ln(0.80) = \left[ \frac{\ln(\mu_T) - \ln(\mu_R)}{2} - \frac{\ln(0.80)}{2} \right] - \left[ \frac{\ln(\mu_R) - \ln(\mu_T)}{2} + \frac{\ln(0.80)}{2} \right].$$

Notons que chaque terme de la dernière formule n'est que l'espérance mathématique de la différence entre l'observation de la période 1 et l'observation de la période 2 du sujet  $i$  ( $i = 1, 2, \dots, n$ ) de la séquence  $k = 1$  et  $k = 2$ , respectivement, de laquelle on soustrait la constante  $\ln(0.80)/2$  pour la séquence 1 et à laquelle on rajoute la constante  $\ln(0.80)/2$  pour la séquence 2. Comme l'espérance mathématique d'une constante est égale à une constante, on peut donc se réduire aux formules suivantes:

$$\begin{cases} \frac{\ln(\mu_T) - \ln(\mu_R) - \ln(0.80)}{2} = \frac{E[y_{i11} - y_{i21}] - \ln(0.80)}{2} = \frac{E[y_{i11} - y_{i21} - \ln(0.80)]}{2}, k = 1, \\ \frac{\ln(\mu_R) - \ln(\mu_T) + \ln(0.80)}{2} = \frac{E[y_{i12} - y_{i22}] + \ln(0.80)}{2} = \frac{E[y_{i12} - y_{i22} + \ln(0.80)]}{2}, k = 2. \end{cases}$$

Cela explique la procédure appliquée aux données des deux séquences mentionnée plus haut.

Par le même raisonnement, pour le test unilatéral d'équivalence (4.2) on soustrait la valeur  $\ln(125)/2$  des observations de la séquence 1 et on rajoute cette valeur aux observations de la séquence 2.

**Étape 4.** On teste l'hypothèse de bioéquivalence par la procédure des deux tests unilatéraux définie par Schuirmann (1981, 1987). On applique le test de permutations, le test  $t$  de Student et le test de Wilcoxon pour deux échantillons sur les données obtenues à l'étape 3. Les traitements Test et Référence sont considérés bioéquivalents, si et seulement si les deux tests unilatéraux rejettent l'hypothèse nulle.

**Étape 5.** On répète les étapes de 1 à 4 plusieurs fois (500 ou 1000 dans notre cas).

Pour réaliser le test de permutations on s'est inspiré de l'article de Balasubramani (2005) qui décrit la procédure de réalisation de toutes les permutations possibles des valeurs des deux échantillons indépendants. Cela diffère de la littérature qui propose d'échantillonner des permutations parmi toutes les permutations possibles.

En effet, le nombre de permutations devient énorme au fur et à mesure que les tailles échantillonales augmentent. Par exemple, pour  $n_1 = n_2 = 5$  le nombre de permutations est de 252, pour  $n_1 = n_2 = 7$  il est de 3432, mais pour  $n_1 = 14$  et  $n_2 = 10$  le nombre de permutations est presque 2,000,000 (voir tableau 4.2).

**Tableau 4.2** Nombre total des permutations pour différentes tailles échantillonales.

Nombre sujets séquence 1 ( $n_1$ )	Nombre sujets séquence 2 ( $n_2$ )	Nombre total de permutations
5	5	252
7	7	3,432
9	9	48,620
14	10	1,961,256

Le temps nécessaire pour réaliser toutes ces permutations croît lui aussi, une fois les tailles échantillonales (qui constituent les séquences dans notre modèle) augmentent. Par exemple, il faut approximativement 310 secondes pour réaliser toutes les 3432 permutations pour des tailles échantillonales  $n_1 = n_2 = 7$  par un ordinateur de vitesse opérationnelle de 1.99 Go. Il ne faut pas oublier aussi qu'on parle ici du nombre de permutations pour la réalisation d'un seul test unilatéral appliqué sur une seule base de données simulée. Cependant, il faut simuler plusieurs bases de données. De plus, pour chacune des bases de données simulée on réalise deux testes unilatéraux, afin d'appliquer la procédure proposée par Schuirman (1981, 1987). Étant donné ces contraintes, on a simulé un total de 500 bases de données de 7 sujets par séquence pour la majorité des exemples analysés dans ce chapitre. Et juste dans un seul cas on a simulé 1000 bases de données (voir tableau 4.4, section des résultats). Le nombre de sujets par séquence est assez réaliste, car en pratique il se peut qu'un minimum de 6 sujets soit recruté par séquence.

Dans la section suivante on présente les résultats obtenus pour les données simulées d'après le modèle du plan croisé 2 x 2, dont les erreurs expérimentales satisfont plusieurs types de loi. Tel que mentionné ci-dessus, les cas sont :

- la loi normale  $\mathcal{N}(0; \sigma_W^2)$  de variance  $\sigma_W^2 = 0.01$  pour certains cas analysés et  $\sigma_W^2 = 0.03$  pour d'autres cas.

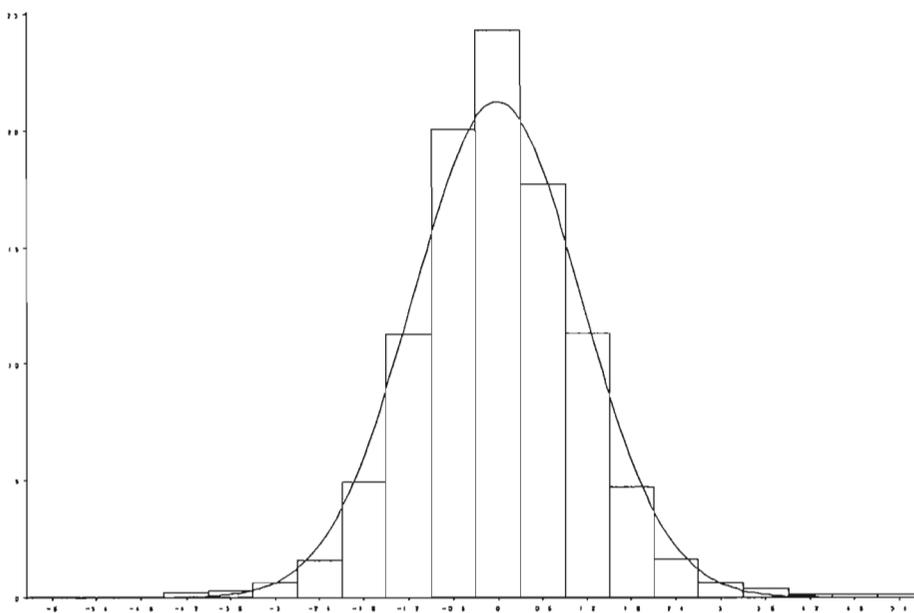
- la loi de mélange composée d'une loi normale  $\mathcal{N}(0, \sigma_W^2)$  et d'une loi double-exponentielle  $\mathcal{DE}(0, \sigma_W/\sqrt{2})$ , les deux de variance  $\sigma_W^2 = 0.03$ .

Dans le cadre de nos analyses on considère deux lois de mélange :

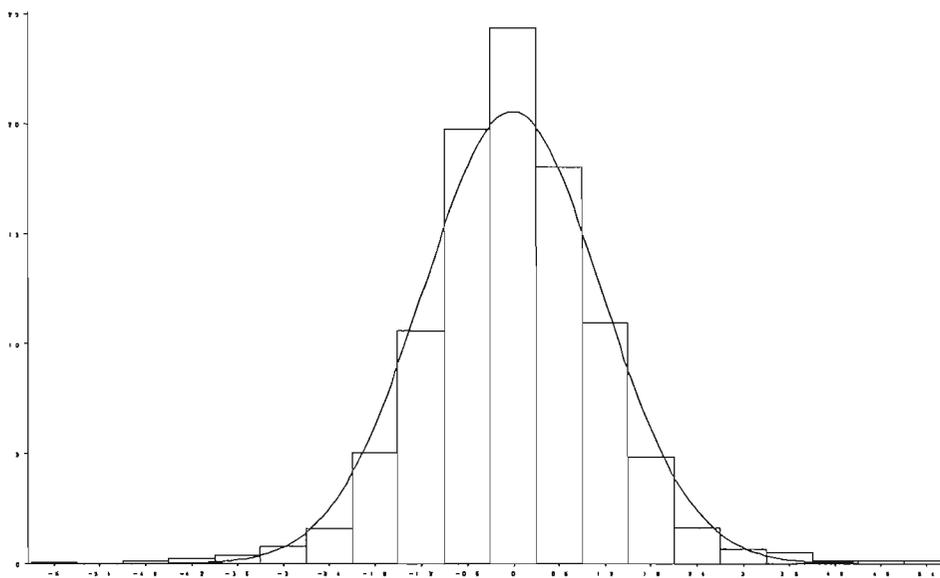
1. la loi composée de 70% loi normale et de 30% loi double-exponentielle ;
2. la loi composée de 60% loi normale et de 40% loi double-exponentielle.

Afin de visualiser ces deux lois, leurs histogrammes obtenus à partir de 3000 simulations des données sont présentés dans les figures 4.1 – 4.2 .

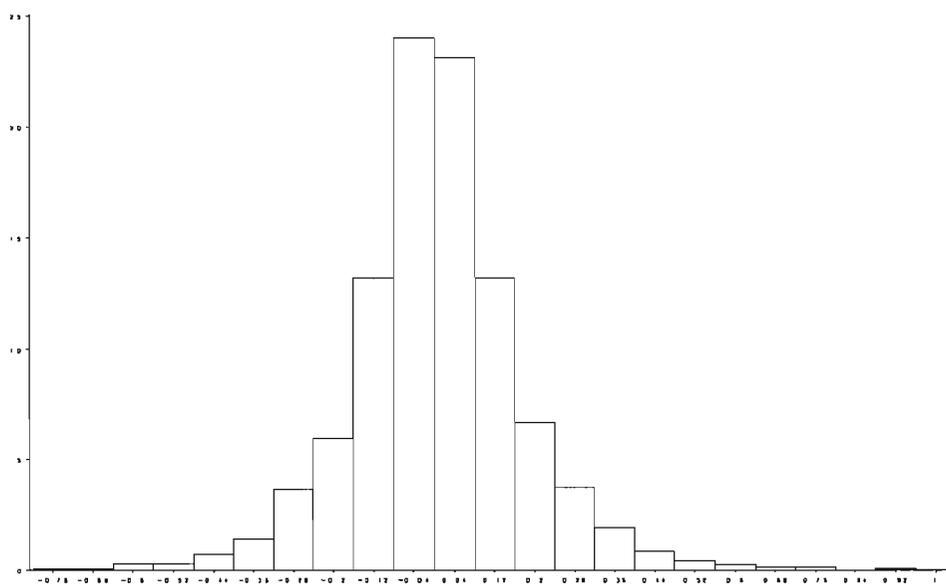
- la loi double-exponentielle  $\mathcal{DE}(0, \sigma_W/\sqrt{2})$  de variance  $\sigma_W^2 = 0.03$  (voir figure 4.3).



**Figure 4.1** Histogramme de loi de mélange composée de 70% loi normale et de 30% loi double-exponentielle et densité de loi normale



**Figure 4.2** Histogramme de loi de mélange composée de 60% loi normale et de 40% loi double-exponentielle et densité de loi normale



**Figure 4.3** Loi double-exponentielle

#### 4.4 Résultats

Les résultats de notre étude sont présentés dans les tableaux 4.3 – 4.7. Dans les tableaux 4.3 – 4.4, sont analysés les cas où les erreurs expérimentales suivent une loi normale. Dans les tableaux 4.5 – 4.6, ces erreurs sont simulées à partir d'un mélange de loi normale et loi double-exponentielle. Et, finalement, dans le tableau 4.7 les erreurs sont simulées à partir d'une loi double-exponentielle. Dans l'analyse on compare principalement le test de permutations avec les deux autres tests.

**Tableau 4.3** Comparaison des valeurs-p (pour un seuil de signification  $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi normale ;  $\sigma_B^2 = 0.015$  ;  $\sigma_W^2 = 0.01$  (500 simulations).

$\mu_T/\mu_R$ (%)	$\ln(\mu_T/\mu_R)$	Test		
		Permutations	T de Student	Wilcoxon
80	-0.223	0.046	0.042	0.044
90	-0.105	0.91	0.91	0.886
100	0	1	1	1
110	0.105	0.936	0.938	0.92
125	0.223	0.038	0.04	0.04

Quand les erreurs expérimentales suivent une loi normale (tableau 4.3) on constate des puissances très élevées (plus que 90%) atteintes par le test de permutations pour chacune des trois valeurs du rapport des moyennes  $\mu_T/\mu_R$  égal à : 90%, 100% et 110%. Ces valeurs sont très comparables à celles obtenues avec le test t de Student. Comparativement aux deux autres tests, le test de Wilcoxon donne les puissances les moins élevées dans ce cas. En ce qui concerne les erreurs de type I, les valeurs obtenues avec le test de permutations sont plus petites que 5% pour les bornes [80%,125%] de l'intervalle de bioéquivalence. Les deux autres tests donnent des erreurs légèrement plus élevées, mais toujours plus petites que 5%,

pour un rapport des moyennes  $\mu_T/\mu_R$  de 125%. Par contre, elles sont moins élevées que les erreurs obtenues avec le test de permutations pour la borne de 80%.

Dans le tableau qui suit (voir tableau 4.4) on a refait nos tests pour les bornes [80%,125%] quand les erreurs expérimentales sont de loi normale, en utilisant plus de simulations (1000 comparativement aux 500 du tableau 4.3).

**Tableau 4.4** Comparaison des valeurs-p (pour un seuil de signification  $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi normale ;  $\sigma_B^2 = 0.015$  ;  $\sigma_W^2 = 0.01$  (1000 simulations).

$\mu_T/\mu_R$ (%)	$\ln(\mu_T/\mu_R)$	Test		
		Permutations	T de Student	Wilcoxon
80	-0.223	0.041	0.038	0.039
125	0.223	0.044	0.044	0.042

Comparativement aux résultats obtenus pour 500 simulations (voir tableau 4.3), les erreurs de type I de la borne supérieure de l'intervalle de bioéquivalence (125%) sont plus élevés pour les trois tests dans le cas où on simule 1000 bases de données (tableau 4.4). Toutefois, elles restent toujours plus petites que 5% et le test de Wilcoxon s'avère le test le plus conservateur. En ce qui concerne les erreurs de type I pour la borne inférieure de l'intervalle de bioéquivalence (80%), la valeur obtenue par le test de permutations est de nouveau la plus élevée (mais plus petite que 5%) parmi les résultats des trois tests. Cependant, on constate une diminution des erreurs de type I dans le cas de 1000 bases simulées comparativement aux 500 bases simulées du cas précédent.

**Tableau 4.5** Comparaison des valeurs-p (pour un seuil de signification  $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi de mélange de loi normale (70%) et de loi double-exponentielle (30%) ;  $\sigma_B^2 = 0.045$  ;  $\sigma_W^2 = 0.03$ .

$\mu_T/\mu_R$ (%)	$\ln(\mu_T/\mu_R)$	Test		
		Permutations	T de Student	Wilcoxon
80	-0.223	<b>0.064</b>	<b>0.064</b>	<b>0.054</b>
90	-0.105	0.52	0.522	0.504
100	0	0.892	0.892	0.874
110	0.105	0.544	0.584	0.546
125	0.223	0.042	0.04	<b>0.052</b>

Pour le mélange composé de la loi normale contaminée avec la loi double-exponentielle dans la proportion de 30% (voir tableau 4.5) on constate qu'il y a une diminution des puissances comparativement aux valeurs du cas de la loi normale. Même si les puissances restent encore assez élevées pour le rapport de moyennes  $\mu_T/\mu_R$  de 100%, elles n'atteignent plus 90%. C'est important de constater à quel point diminuent les puissances calculées lorsque le rapport des moyennes  $\mu_T/\mu_R$  est égal à 90% et 110%. Ces valeurs sont un peu plus élevées que 0.5. Comparativement aux résultats obtenus pour les mêmes valeurs du rapport des moyennes  $\mu_T/\mu_R$  pour la loi normale, ces puissances sont presque 1.8 fois moins élevées. Donc, pour un rapport des moyennes  $\mu_T/\mu_R$  égal à 90% et 110%, on rejettera environ une fois sur deux l'hypothèse alternative de bioéquivalence quand en réalité les deux traitements étudiés sont bioéquivalents. Cette tendance est la même à travers les trois tests.

Les puissances obtenues avec le test de permutations sont très comparables à celles du test t de Student, tandis que le test de Wilcoxon donne des valeurs plus petites.

On constate une erreur de type I plus élevée que 5% pour la borne de 80% (voir valeurs en gras, Tableau 4.5) et une erreur moins élevée que 5% pour la borne de 125% dans le cas

du test de permutations. Les erreurs de type I du test t de Student sont comparables à celles obtenues par le test de permutations. Par contre, le test de Wilcoxon donne une erreur moins élevée que les autres deux tests pour la borne de 80% et une erreur plus élevée pour la borne de 125% (voir valeurs en gras, Tableau 4.5). Rappelons-nous que les hypothèses du test de bioéquivalence sont inversées par rapport à un test d'hypothèses ordinaire. Et cela dans le but de pouvoir contrôler le risque d'accepter l'hypothèse de bioéquivalence dans le cas où les deux traitements sont bioéquivalents. Dans ce contexte, avec le test de Wilcoxon on acceptera moins de fois la bioéquivalence des traitements quand cette hypothèse est fausse qu'avec les deux autres tests pour la borne inférieure de l'intervalle de bioéquivalence (80%). Tandis que pour la borne supérieure de bioéquivalence (125%), le risque de prendre une fausse décision en faveur de la bioéquivalence est un peu plus élevé que 5% pour le test de Wilcoxon.

Dans le cas suivant (voir tableau 4.6) on refait nos tests pour la loi de mélange, mais on considère déjà une contamination plus élevée de la loi normale par la loi double-exponentielle (60% loi normale et 40% loi double-exponentielle) que dans le cas qu'on vient d'analyser.

**Tableau 4.6** Comparaison des valeurs-p (pour un seuil de signification  $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi de mélange de loi normale (60%) et de loi double-exponentielle (40%) ;  $\sigma_B^2 = 0.045$  ;  $\sigma_W^2 = 0.03$ .

$\mu_T/\mu_R$ (%)	$\ln(\mu_T/\mu_R)$	Test		
		Permutations	T de Student	Wilcoxon
80	-0.223	0.034	0.032	0.028
90	-0.105	0.516	0.514	0.488
100	0	0.858	0.864	0.822
110	0.105	0.568	0.576	0.562
125	0.223	<b>0.058</b>	<b>0.058</b>	<b>0.06</b>

Malgré une diminution plus importante des puissances obtenues par les trois tests (voir tableau 4.6), les conclusions restent essentiellement les mêmes que dans le cas précédent (voir tableau 4.5). Par contre, les erreurs de type I sont moins élevées pour le rapport  $\mu_T/\mu_R$  de 80% et plus élevées pour le rapport  $\mu_T/\mu_R$  de 125%. Comme ces dernières erreurs sont plus grandes que 5% (voir valeurs en gras, Tableau 4.6), le risque de prendre une fausse décision en faveur de la bioéquivalence est plus élevé pour les trois tests dans ce cas.

**Tableau 4.7** Comparaison des valeurs-p (pour un seuil de signification  $\alpha = 5\%$ ) des trois tests quand les erreurs expérimentales suivent une loi double-exponentielle;

$$\sigma_B^2 = 0.045 ; \sigma_W^2 = 0.03.$$

$\mu_T/\mu_R$ (%)	$\ln(\mu_T/\mu_R)$	Test		
		Permutations	T de Student	Wilcoxon
80	-0.223	0.046	0.046	0.042
90	-0.105	0.56	0.556	0.558
100	0	0.858	0.858	0.868
110	0.105	0.59	0.59	0.584
125	0.223	<b>0.054</b>	<b>0.056</b>	<b>0.052</b>

Pour la loi double-exponentielle (voir tableau 4.7), on constate que les puissances obtenues avec le test de permutations pour un rapport des moyennes  $\mu_T/\mu_R$  de 100% sont semblables aux puissances obtenues pour la loi de mélange composée de 60% de loi normale et de 40% de loi double-exponentielle. Quant aux puissances pour les valeurs du rapport  $\mu_T/\mu_R$  égales à 90% et 110% on constate le même phénomène que dans le cas des lois de mélange. Ces puissances ne dépassent pas 60% et, donc, on peut dire qu'on rejettera l'hypothèse alternative de bioéquivalence en environ 40% des fois où en réalité les deux traitements étudiés sont bioéquivalents.

Comme dans les cas précédents, les puissances obtenues avec le test de permutations sont très comparables à celles évaluées par le test t de Student. Cependant, ce type de loi rentre dans la catégorie de celles où le test de Wilcoxon commence à mieux performer que le test de t de Student, par exemple. Pour un rapport  $\mu_T/\mu_R$  de 100% le test de Wilcoxon est légèrement plus puissant que les autres tests. De plus, ce dernier s'avère le test le plus conservateur parmi les trois tests, car pour cette loi les erreurs de type I sont moins élevées que celles obtenues par le test de permutations et le test t de Student. Comme les trois tests dépassent le seuil de  $\alpha = 5\%$  (voir valeurs en gras, Tableau 4.7), le test de Wilcoxon est à nouveau le meilleur des trois pour la valeur de 125%, suivi par le test de permutations.

Pour résumer, on peut dire que pour tous les cas analysés on constate une puissance plus élevée que 85% pour le test de permutations quand le rapport  $\mu_T/\mu_R$  est égal à 100% et ces résultats sont très comparables à ceux obtenus par les deux autres tests. Même dans les cas où les erreurs expérimentales ne suivent pas des lois normales (cas de la loi de mélange et de la loi double-exponentielle) les puissances obtenues pour toutes les valeurs étudiées du rapport  $\mu_T/\mu_R$  par le test t de Student sont très proches, voir égales, à celles obtenues par le test de permutations. Le test de Wilcoxon s'avère le moins puissant dans le cas de la loi normale et de la loi de mélange, mais il donne une puissance légèrement plus élevée que les autres deux tests pour la loi double-exponentielle pour un rapport  $\mu_T/\mu_R$  de 100%.

Lorsque le rapport des moyennes  $\mu_T/\mu_R$  est égal à 90% et à 110% et les erreurs expérimentales suivent une loi de mélange ou une loi double-exponentielle, les puissances obtenues par les trois tests sont autour de 0.5. Donc, dans ces cas, on rejettera environ une fois sur deux l'hypothèse alternative de bioéquivalence quand en réalité les deux traitements étudiés sont bioéquivalents. Cette constatation est particulièrement intéressante pour la pratique médicale.

On a donc constaté que le test de permutations est très puissant et il donne des valeurs très semblables à celles obtenues par le test t de Student pour la loi normale et les lois mixtes définies dans notre étude (mélanges de loi normale et double-exponentielle). Ces puissances sont plus élevées que celles du test de Wilcoxon pour les lois énumérées. Par contre, la loi double-exponentielle rentre dans la catégorie des lois où le test de Wilcoxon commence à

mieux performer que les autres deux tests et dans ce dernier cas la puissance du test de Wilcoxon est légèrement plus élevée que celle du test de permutations et du test t de Student. En analysant les erreurs de type I obtenues dans notre étude de simulation, on a retrouvé un résultat attendu, que, dans presque tous les cas analysés, le test de Wilcoxon est plus conservateur que les deux autres tests. Par contre, lorsque les données ne sont plus normales, même les deux autres tests peuvent dépasser légèrement la probabilité de 5% pour un rapport situé à la frontière, ce qui veut dire qu'on risque d'accepter la bioéquivalence quand elle n'est pas vraie, même si ce risque est assez petit.

Toutefois, on pense que, si on avait pu réaliser beaucoup plus de simulations, on aurait pu dégager des conclusions plus tranchantes pour tous les tests et, en particulier, pour le test de permutations. D'après la littérature, ce test est plus puissant si on le compare aux tests purement non paramétriques et, de plus, il assure le contrôle de l'erreur de type I sous une condition très minimale: de randomiser les unités expérimentales (les sujets dans le cas d'une étude de bioéquivalence) aux traitements (Ernst, 2004). En effet, avec la procédure de Balasubramani (2005) le nombre de permutations devient énorme au fur et à mesure que les tailles échantillonnales augmentent, ce qui requiert beaucoup de temps de calcul. Une autre approche de la littérature récente est d'échantillonner des permutations parmi toutes les permutations possibles. En appliquant cette procédure, le temps de réalisation des tests de permutation est diminué considérablement.

## CONCLUSION

Tel que mentionné dans ce mémoire, une étude de bioéquivalence consiste à comparer un nouveau traitement, Test, à un traitement standard, Référence, et à estimer leur équivalence en termes de mesures de biodisponibilité médicale. Généralement, il existe trois types d'études de bioéquivalence: moyenne, de population et individuelle. Dans ce travail, nous avons étudié la bioéquivalence moyenne qui compare les moyennes des traitements,  $\mu_T$  et  $\mu_R$ .

Deux éléments caractérisent les études de bioéquivalence. En premier, pour contrôler le risque de prendre une mauvaise décision en faveur de l'hypothèse d'équivalence des traitements, on inverse l'hypothèse nulle et l'hypothèse alternative. Ainsi, on prend l'inéquivalence des traitements Test et Référence comme hypothèse nulle et l'équivalence des traitements comme hypothèse alternative. De plus, l'hypothèse alternative d'une étude de bioéquivalence est définie par un intervalle d'équivalence. Les limites de cet intervalle sont connues et établies d'avance.

Après avoir présenté les bases théoriques du concept de bioéquivalence et les analyses statistiques les plus courantes, nous avons effectué une étude de simulation afin de comparer plusieurs tests statistiques de la pratique biomédicale, auxquels on a ajouté un test de permutations. Ainsi, à travers notre étude de simulation, on a comparé la performance du test de permutations par rapport au test t de Student et au test de Wilcoxon sur des données qui proviennent de plusieurs types de lois. Le test de permutations a été réalisé selon la procédure proposée par Balasubramani (2005) qui effectue toutes les permutations possibles des valeurs des deux échantillons indépendants, une procédure qui prend beaucoup de temps de calcul.

En général, on a constaté que le test de permutations est très puissant et il donne des valeurs très semblables à celles obtenues par le test t de Student pour la loi normale et les lois mixtes définies dans notre étude (mélanges de loi normale et double-exponentielle). Ces

puissances sont plus élevées que celles du test de Wilcoxon pour les lois énumérées. Par contre, la loi double-exponentielle rentre dans la catégorie des lois où le test de Wilcoxon commence à mieux performer que les autres deux tests et dans ce dernier cas la puissance du test de Wilcoxon est légèrement plus élevée que celle du test de permutations et du test t de Student. En analysant les erreurs de type I obtenues dans notre étude de simulation, on a retrouvé un résultat attendu, que, dans presque tous les cas analysés, le test de Wilcoxon est plus conservateur que les deux autres tests. Par contre, lorsque les données ne sont plus normales, même les deux autres tests peuvent dépasser légèrement la probabilité de 5% pour un rapport situé à la frontière, ce qui veut dire qu'on risque d'accepter la bioéquivalence quand elle n'est pas vraie, même si ce risque est assez petit. Toutefois, on pense que, si on avait pu réaliser beaucoup plus de simulations, on aurait pu dégager des conclusions plus tranchantes pour tous les tests et, en particulier, pour le test de permutations.

Notre plus importante constatation est la baisse importante de la performance de tous les tests quand les données sont contaminées par des ailes non normales et on s'éloigne du centre de l'intervalle de bioéquivalence (cas de l'égalité parfaite), par exemple lorsqu'on atteint un rapport  $\mu_T/\mu_R$  de 90% ou 110%. Nous avons trouvé des puissances d'environ 55% dans la plupart de ces cas. Ce résultat est intéressant pour la pratique, où on utilise depuis longtemps des tests non paramétriques afin de tenir compte de la possibilité que les données ne soient pas normales. Ce qui se passe, malgré tout, est qu'environ la moitié du temps on ne peut pas détecter la bioéquivalence, si les deux moyennes ne sont pas suffisamment rapprochées.

APPENDICE A

RÈGLE TRAPÉZOÏDALE

A.1 Règle trapézoïdale d'approximation de l'intégrale  $\int_a^b f(x)dx$  ..... 113

### A.1 Règle trapézoïdale d'approximation de l'intégrale $\int_a^b f(x)dx$

Soit  $f(x)$  une courbe continue définie sur l'intervalle  $[a, b]$ . La méthode trapézoïdale consiste à diviser l'intervalle  $[a, b]$  en sous-intervalles  $a = x_0 < x_1 < \dots < x_n = b$ . La surface sous la courbe peut être estimée en utilisant l'approximation trapézoïdale sur chacun des sous-intervalles  $[x_{i-1}, x_i]$ . En joignant les points  $(x_i, f(x_i))$  par des lignes droites, on obtient un ensemble des trapèzes. Leur surface peut être calculée en utilisant la formule

$$A_i = 1/2 [f(x_{i-1}) + f(x_i)] \Delta x_i$$

(une demi de la somme des bases du trapèze multiplié par sa hauteur).

L'approximation de l'intégrale  $\int_a^b f(x)dx$  par la règle trapézoïdale est définie par la formule

$$\sum_{i=1}^n 1/2 [f(x_{i-1}) + f(x_i)] \Delta x_i. \quad (\text{A.1})$$

Généralement les points  $x_i$  sont équidistants, où  $\Delta x_i = (b - a)/n$ ,  $x_i = a + i(b - a)/n$ . Dans ces conditions, la formule (A.1) peut être écrite sous la forme

$$((b - a)/n) \sum_{i=1}^n 1/2 [f(x_{i-1}) + f(x_i)] \quad (\text{A.2})$$

## APPENDICE B

### MACRO SAS

**B.1 Macro de simulation des données d'après le modèle du plan croisé 2 x 2 avec des effets aléatoires  $S_{ik}$ , ( $i = 1, 2, \dots, n$  ;  $k = 1, 2$ ) et des erreurs expérimentales  $e_{ijk}$ , ( $i = 1, 2, \dots, n$  ;  $k = 1, 2$ ;  $j = 1, 2$ ) pour la loi normale (Patterson et Jones, 2006)..... | 15**

B.1 Macro de simulation des données d'après le modèle du plan croisé 2 x 2 avec des effets aléatoires  $S_{ik}$ , ( $i = 1, 2, \dots, n$ ;  $k = 1, 2$ ) et des erreurs expérimentales  $e_{ijk}$ , ( $i = 1, 2, \dots, n$ ;  $k = 1, 2$ ;  $j = 1, 2$ ) pour la loi normale (Patterson et Jones, 2006).

```
%macro simul_data(seed1,seed2,seed3,seed4,seed5,seed6,valuT, valmuR, param);
data simulTR simulRT;
per1=0; /*effet période 1*/
per2=0; /*effet période 2*/
lamdaT=0; /*effet résidu du traitement Test*/
lamdaR=0; /*effet résidu du traitement Référence*/
varB=0.015; /*variabilité inter-sujet*/
varW=0.01; /*variabilité intra-sujet*/
muT=&valuT; /*la moyenne des logarithmes naturel des valeurs Test */
muR=&valmuR; /*la moyenne des logarithmes naturel des valeurs Référence */

/*Simulation des données pour la séquence 1*/
do i=1 to 13 by 2;
E11=rannor(&seed1);
E12=rannor(&seed2);
S1=rannor(&seed3);
y11=per1+sqrt(varB)*S1+muT+sqrt(varW)*E11;
y12=per2+sqrt(varB)*S1+muR+lamdaT+sqrt(varW)*E12;
subj=i;
seq=1;
output simulTR;
end;

/*Simulation des données pour la séquence 2*/
do i=2 to 14 by 2;
E21=rannor(&seed4);
E22=rannor(&seed5);
S2=rannor(&seed6);
y21=per1+sqrt(varB)*S2+muR+sqrt(varW)*E21;
y22=per2+sqrt(varB)*S2+muT+lamdaR+sqrt(varW)*E22;
subj=i;
seq=2;
output simulRT;
end;

run;
data simulTR;
set simulTR;
keep y11 y12 seq subj;
run;

data simulRT;
set simulRT;
keep y21 y22 seq subj;
run;

data simulTR;
set simulTR;
```

```
rename y11=T;
rename y12=R;
run;

data simulRT;
set simulRT;
rename y21=R;
rename y22=T;
run;

data simul (keep=subj seq T R);
set simulTR simulRT;
run;

proc sort data=simul;
by subj seq;
run;

proc transpose data=simul out=simul_t prefix=&param;
by subj seq;
var T R;
run;

data simul_out;
set simul_t;
treat=_name_;
rename &param.1=&param;
if seq=1 and treat='T' then per=1;
if seq=1 and treat='R' then per=2;
if seq=2 and treat='R' then per=1;
if seq=2 and treat='T' then per=2;
drop _name_;
run;

%mend simul_data;
```

## BIBLIOGRAPHIE

- Balasubramani, G.K., Stephen R. Wisniewski, Hongwei Zhang et Heather F. Eng, «Development of an efficient SAS macro to perform permutations tests for two independent samples», *Computer Methodes and Programs in Biomedicine*, Pittsburg, USA, 2005.
- Chow, Shein-Chung et Jen-Pei Liu, *Design and Analysis of Bioavailability and Bioequivalence studies*, 2000.
- DiPiro, Joseph T., et al, *Concepts in Clinical Pharmacology*, American Society of Health-System Pharmacists, 2005.
- Ernst, Michael D., «Permutation Methodes: A Basis for Exact Inference», *Statistical Science*, Vol.19, No.4, 676-685, 2004.
- Gibaldi, Milo, et Donald Perrier, *Pharmacokinetics*, 2<sup>d</sup> edition, revised and expanded, New York: Dekker, 1982.
- Hauschke, D., et V.W. Steijans, E. Diletti, «A distribution-free procedure for the statistical analysis of bioequivalence studies», *International Journal of Clinical Pharmacology, Therapy and Toxicology*, Vol.28, No.2, 72-78, 1990.
- Koch, Gary G., «The Use of Non-Parametric Methodes in the Statistical Analysis of the Two-Period Change-Over Design», *Biometrics*, Vol.28, No.2., 577-584, 1972.
- Kuehl, Robert O., *Design of Experiments : Statistical Principles of Research Design and Analysis*, 2000.
- Patterson, Scott et Jones, Byron, *Bioequivalence and Statistics in Clinical Pharmacology*, 2006.
- Putt, Mary E. et Chinchilli, Vernon M., «Nonparametric Approaches to the Analysis of Crossover Studies», *Statistical Science*, Vol.19, No.4, 712-719, 2004.
- Schuurmann, DJ, «A comparison of the two one-sided tests procedure and power approach for assessing the equivalence of average bioavailability», *Journal of Pharmacokinetics and Biopharmaceutics*, Vol.15, No.6, 657-680, 1987.
- Hauck, Walter W. et all, «A Parametric Approach to Population Bioequivalence», *Statistics in Medicine*, Vol.16, 441-454, 1997.

Wellek, Stefan, *Testing Statistical Hypotheses of Equivalence*, 2003.