



# Transfert d'un test biologique en développement clinique : évaluation d'une nouvelle approche statistique

Baptiste Klehammer

## ► To cite this version:

Baptiste Klehammer. Transfert d'un test biologique en développement clinique : évaluation d'une nouvelle approche statistique. Méthodologie [stat.ME]. 2014. dumas-01059939

**HAL Id: dumas-01059939**

**<https://dumas.ccsd.cnrs.fr/dumas-01059939>**

Submitted on 9 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GlaxoSmithKline Vaccines  
Vaccines Discovery & Development  
Rue de l'Institut 89  
1330 Rixensart, Belgique



Université de Strasbourg  
U.F.R. de Mathématiques  
7, Rue René Descartes  
67084 Strasbourg, France



# **Rapport de stage**

## **Master de Statistique - 2ème année**

### **10 Février 2014 - 31 Juillet 2014**

**Intitulé : Transfert d'un test biologique en développement clinique  
Evaluation d'une nouvelle approche statistique**

**KLEHAMMER Baptiste**

**Stagiaire : KLEHAMMER Baptiste**  
**Maître de Stage : DEWE Walthère**

# Table des matières

<b>1. Présentation de GlaxoSmithKline (GSK)</b>	<b>3</b>
1.1. Un leader mondial . . . . .	3
1.2. GSK en Belgique . . . . .	3
<b>2. Familiarisation avec le milieu</b>	<b>4</b>
<b>3. Introduction au bridging</b>	<b>5</b>
<b>4. Bridging : partie clinique</b>	<b>7</b>
4.1. Méthodologie . . . . .	7
4.2. Simulation . . . . .	7
4.3. Intérêt de la comparaison anciens/nouveaux . . . . .	8
4.4. Simplification par développement analytique et contraintes . . . . .	9
4.5. Exemples . . . . .	10
4.6. Conclusion . . . . .	11
<b>5. Bridging : partie technique</b>	<b>13</b>
5.1. Présentation . . . . .	13
5.2. Régression et modèles errors-in-variables . . . . .	14
5.3. Intervalles de prédiction . . . . .	14
<b>6. Conclusion</b>	<b>16</b>
<b>Bibliographie</b>	<b>16</b>
<b>A. Simulations</b>	<b>18</b>
<b>B. Approche analytique</b>	<b>19</b>

# **1. Présentation de GlaxoSmithKline (GSK)**

## **1.1. Un leader mondial**

GlaxoSmithKline est une entreprise britannique, numéro 6 de l'industrie pharmaceutique mondiale en 2013. Elle résulte de la fusion entre Glaxo Wellcome et SmithKline Beecham en 2000. GSK possède trois branches d'activité : le laboratoire GlaxoSmithKline (constitué de GSK Pharmaceuticals et GSK Vaccines) et GlaxoSmithKline Santé Grand Public. Le laboratoire GlaxoSmithKline est présent dans de nombreux domaines thérapeutiques : la neurologie, la pneumologie, le diabète, l'infection par le VIH, etc. et occupe une place prépondérante en vaccinologie. GlaxoSmithKline Santé Grand Public est impliqué dans la dermatologie, l'hygiène bucco-dentaire et l'automédication. Aujourd'hui, GSK emploie près de 100 000 collaborateurs dans plus de 100 pays, et vend plus de 100 médicaments de prescription et de vaccins dans 140 pays.

## **1.2. GSK en Belgique**

GSK Vaccines, la division vaccins de GlaxoSmithKline, est un leader mondial dans la production de vaccins. Cette division est basée à Rixensart, site où j'ai effectué mon stage, et à Wavre en Belgique où sont réalisées la majorité des activités de GlaxoSmithKline en matière de vaccinologie. GSK Vaccines emploie plus de 1 600 scientifiques, qui se consacrent à la découverte de nouveaux vaccins et au développement de produits moins chers et plus faciles d'utilisation, afin de prévenir des infections qui sont la cause de sérieux problèmes médicaux dans le monde. GSK Vaccines distribue chaque année plus de 1,2 milliard de doses de vaccins dans 173 pays, soit une moyenne de 35 doses toutes les secondes.

Le site de Rixensart emploie quotidiennement près de 2000 personnes dans une cinquantaine de bâtiments et se consacre à la recherche, développement et production des vaccins. J'ai commencé mon stage dans la section Global Clinicals Operations du département Clinical Statistics BE avant de suivre mon maître de stage dans la section Shared Services Vaccines Discovery & Development du département Biostatistics et j'ai à cette occasion pu travailler avec deux équipes différentes : celle de Marc Fourneau puis celle de Maria Key Prato. La première équipe se concentrait sur le développement clinique relatif aux vaccins alors que la seconde était axée sur la partie non-clinique et couvrait les activités de recherche et développement.

## 2. Familiarisation avec le milieu

Le développement d'un vaccin est un processus long et coûteux (9 à 14 ans pour environ 500 millions d'euros) qui peut se résumer en 7 phases :

- Phase exploratoire : Identification des antigènes pour la sélection de candidats vaccins qui poursuivront le processus
- Phase préclinique : Evaluation des antigènes sur l'animal et sélection du meilleur candidat vaccin
- Phase I : Test d'innocuité sur 10 à 100 sujets
- Phase II : Réponse immunitaire sur 100 à 3 000 sujets
- Phase III : Etude d'efficacité, de tolérance à grande échelle et de consistance sur 3 000 à 40 000 sujets
- Enregistrement du produit
- Lancement du produit

Sans trop entrer dans les détails, voici une petite mise en contexte permettant de comprendre certains termes techniques utilisés et de brièvement décrire le fonctionnement d'un vaccin. Le système immunitaire est l'ensemble des mécanismes chargés de la défense de notre organisme au quotidien, face aux agressions extérieures. Cette protection naturelle est aussi activée par l'injection d'un vaccin, qui permet de provoquer une réaction immunitaire au sein de l'organisme. Celle-ci aboutit à la mise en place d'une défense qui peut être de deux types : humorale ou cellulaire. L'immunité humorale entraîne la production d'anticorps (réponse à une stimulation par un antigène) par l'organisme, pour répondre à l'introduction d'un antigène (substance étrangère au corps capable de provoquer une réponse du système immunitaire), tandis que l'immunité cellulaire suscite une modification des cellules qui sont chargées de détruire les organismes étrangers.

### 3. Introduction au bridging

Au cours du développement d'un vaccin, différents tests cliniques sont réalisés et il arrive que pour un test donné, les conditions changent. Cela peut se traduire par un changement d'opérateur ou de laboratoire par exemple, et cela influe sur les mesures relevées. Le principe du bridging est de déterminer dans quelles proportions ces mesures peuvent varier afin de ne pas altérer les conclusions cliniques de l'étude. Par extension, il s'agit de s'assurer qu'un changement de condition n'influera pas de manière significative la conclusion de l'étude.

Le bridging se découpe en deux parties : une partie technique qui se focalise sur la méthode analytique et une partie clinique. La méthode actuelle du bridging clinique lors d'un changement dans la manière de mesurer consiste à retester tous les échantillons afin de comparer les deux manières de mesurer sur les mêmes échantillons. Cela n'est non seulement pas toujours possible (les échantillons ne sont plus disponibles etc...) mais requiert également du temps, de l'argent et du matériel biologique. C'est pour cela qu'une nouvelle méthode a été proposée, à base de simulations, et sur laquelle je me suis concentrée. Le but de cette approche est de trouver des spécifications pour le bridging technique. Chronologiquement, j'ai d'abord été assigné à la partie clinique avant de rejoindre un groupe de travail pour la partie technique.

Les études sur lesquelles un bridging est réalisé sont constituées d'une comparaison entre un groupe Traitement et un groupe Contrôle, et dont les données brutes sont des titres (censures par intervalles) ou des concentrations. Le titrage est une méthode de mesure d'une concentration utilisée lorsque l'on ne peut pas la mesurer directement et qui consiste grossièrement à effectuer une dilution de la solution titrée avec une solution titrante par seuils jusqu'à observer un changement de couleur, de potentiel ou du pH, ce qui permet d'obtenir un intervalle contenant la vraie concentration. La comparaison porte généralement sur les moyennes géométriques des titres (GMT) et le taux de séroprotection (SPR) ou séroconversion (SCR). Le taux de séroprotection est la proportion de sujets dont le titre d'anticorps post-vaccination dépasse un certain seuil, identifié comme protecteur, tandis que le taux de séroconversion est généralement la proportion de sujets dont le titre d'anticorps a suffisamment augmenté (quadruplé) entre la mesure pré-vaccination et la mesure post-vaccination<sup>[5]</sup>.

Les objectifs initiaux du stage sont axés sur les trois points suivants :

- Comprendre et revoir ce qui a été fait.
- Proposer des améliorations.

- Ecrire et présenter un compte-rendu.

Lors du bridging, certaines hypothèses de travail sont faites et en voici une liste exhaustive :

- les données suivent une loi log-normale<sup>[5]</sup>
- les erreurs de mesure peuvent varier de 10 à 30% selon les cas
- les erreurs de mesure suivent une loi normale<sup>[5]</sup>
- le rapport des coefficients de variation des erreurs de mesure entre deux jeux de mesures (opérateurs ou laboratoires différents) est constant
- il existe une relation linéaire entre les logarithmes des anciennes et nouvelles valeurs dans la gamme des mesures considérées

## 4. Bridging : partie clinique

### 4.1. Méthodologie

Après m'être rapidement familiarisé avec mon environnement de travail, je me suis attelé à l'analyse d'un rapport sur la nouvelle méthode de bridging clinique, qui était alors à l'état de prototype. Ce rapport a été sous-traité et présentait des résultats surprenants. Dans l'étude qui a servi de base à ce rapport, les données brutes étaient les titres d'anticorps post-vaccination de deux groupes et une comparaison de non-infériorité était réalisée sur les GMR et SPR. Le bridging clinique était réalisé de la façon suivante : pour chaque couple d'ordonnée à l'origine  $\alpha$  et de pente  $\beta$ , on simule des nouvelles mesures (notées Y ici) puis on effectue un test de non-infériorité sur les GMR et SPR ainsi qu'une comparaison entre les anciennes mesures et les nouvelles à l'aide d'un test d'équivalence sur les moyennes géométriques.

Ainsi les mesures simulées (Y) étaient reliées aux anciennes mesures (notées X ici) à l'aide de la relation suivante :  $\log_{10}(Y) = \alpha + \beta \cdot \log_{10}(X) + \epsilon$  où  $\epsilon$  représentait l'erreur de mesure. On garde ensuite les couples de valeurs pour lesquelles les trois tests précédents sont acceptés simultanément pour au moins 90% des jeux de données simulés. On obtient finalement une grille de valeurs pour lesquelles le bridging (du moins pour les deux critères évoqués) est réussi, i.e. que la conclusion de l'étude clinique reste inchangée. Malheureusement, la zone obtenue, que cette nouvelle approche cherchait à trouver, semblait être trop restreinte et c'est sur ce point que je me suis concentré en priorité.

### 4.2. Simulation

La première étape dans la reproduction des résultats précédents consiste à choisir une méthode de simulation pour les jeux de données. Deux approches sont possibles.

L'approche appelée « par perturbation » consiste à faire varier les anciennes valeurs de manière à reproduire des données similaires avec des erreurs de mesure différentes mais dont l'ordre de grandeur est le même. En pratique, chaque ancienne valeur est multipliée par un coefficient proche de 1.

L'approche alternative est basée sur la méthode de la déconvolution et permet d'estimer une densité à partir d'un échantillon contenant des erreurs de mesure<sup>[7,8]</sup>. Une fois la densité réelle estimée, on peut tirer des échantillons dans cette densité et y « ajouter » des erreurs de mesure. Cette approche a été



écartée car trop coûteuse en temps de calcul pour un bénéfice qui ne semble pas réel.

Plus de détails concernant la première méthode sont disponibles en appendice.

On note  $X_{i,old}, i \in [1; n]$  les anciennes mesures des titres et  $X_i, i \in [1; n]$  les nouvelles mesures simulées où  $n$  représente le nombre total (pré + post pour chaque groupe) de mesures effectuées et gardées. En notant  $Y_i, i \in [1; n]$  les mesures simulées avec un intercept  $\alpha$  et une pente  $\beta$  par l'hypothèse de linéarité, on utilise les relations suivantes dans la simulation « par perturbation » :

$$\log_{10}(X_i) = \log_{10}(X_{i,old}) + \epsilon, \epsilon \sim \mathcal{N}(0, 0.0860)$$

$$\log_{10}(Y_i) = \alpha + \beta \cdot \log_{10}(X_i) + \epsilon, \epsilon \sim \mathcal{N}(0, 0.0860)$$

Avant de lancer les simulations sur une large grille de couples de valeurs, il faut non seulement s'assurer de l'exactitude des calculs effectués mais aussi de leur rapidité d'exécution, et choisir un compromis si nécessaire. C'est notamment le cas du calcul des intervalles de confiance d'une différence de proportions où plusieurs choix sont possibles : l'intervalle de Wald, l'intervalle d'Agresti-Caffo et l'intervalle du score<sup>[2]</sup>. Des simulations ont été effectuées afin d'estimer les probabilités de couverture et les temps de calculs des différentes méthodes. Les résultats sont les suivants :

- Les méthodes d'Agresti-Caffo et du score ont une probabilité de couverture plus proche (et supérieure) du niveau nominal (95 %) que la méthode de Wald
- Les méthodes d'Agresti-Caffo et de Wald sont environ 50 fois plus rapides en temps de calcul que la méthode du score

Etant donné que la différence de temps de calcul était significative (0.05 s vs 2.50 s) et que les méthodes d'Agresti-Caffo et du score avaient une probabilité de couverture similaire, la méthode d'Agresti-Caffo a été retenue pour calculer les intervalles de confiance d'une différence de proportions lors des simulations.

Par ailleurs, R n'étant par défaut pas capable d'exploiter toute la puissance de calcul d'un processeur à plusieurs coeurs, et comme la simulation est principalement constituée de boucles indépendantes, j'ai implémenté une parallélisation des calculs et approximativement diviser le temps des simulations par le nombre de coeurs disponibles sur l'ordinateur.

### 4.3. Intérêt de la comparaison anciens/nouveaux

Parmi les différents tests réalisés, la comparaison effectuée entre les anciennes valeurs et les nouvelles (à l'aide des moyennes géométriques) n'était

pas justifiée et je me suis donc penché sur la pertinence de cette comparaison. Si cette comparaison est a priori sensée, on s'aperçoit vite que ce n'est pas le cas et qu'elle n'a pas lieu d'être. En effet, les limites d'acceptation que l'on fixe impactent directement la régression de Deming qui est faite sur les données lors du bridging dans le sens où des limites d'acceptation restreintes donneront des intervalles restreints pour l'ordonnée à l'origine et la pente lors du bridging. En d'autres termes, cette comparaison est redondante et doit absolument être supprimée.

Par ailleurs, en essayant de déterminer comment chacun des tests effectués influait la zone d'acceptation (zone verte dans les deux exemples décrits dans les parties suivantes), je me suis aperçu que pour l'étude sur laquelle portait le bridging clinique qui m'était présenté, la comparaison entre les anciennes et nouvelles valeurs était la seule responsable d'une zone d'acceptation si étroite.

#### 4.4. Simplification par développement analytique et contraintes

En travaillant sur une approche analytique du problème, on arrive à mettre des spécifications sur la pente dans certains cas. En effet, la relation log-linéaire qui existe entre les nouvelles et les anciennes valeurs peut être utilisée afin d'isoler la pente (cf. Approche analytique). Il sera intéressant de vérifier que les résultats ultérieurement obtenus par simulations concordent avec ces spécifications. Par ailleurs, il n'est pas possible de trouver de manière similaire de spécifications sur l'ordonnée à l'origine : ces dernières seront décidées à l'aide des simulations.

Au cours des premières simulations, le bridging était validé pour des pentes négatives, ce qui était assez déroutant mais pas incompréhensible. En effet, pour une pente égale à -1 (et une ordonnée à l'origine égale à 0) par exemple, on a  $GMR_Y = 1/GMR_X$  ce qui implique que le test effectué sur  $GMR_Y$  aura le même résultat que celui effectué sur  $GMR_X$  puisque  $IC(GMR_X^\beta) = IC^\beta(GMR_X)$  (les détails sur les notations et calculs sont disponibles en appendice). De même, les proportions calculées resteront inchangées, ce qui fait que le bridging sera validé sur ces deux critères alors que cette situation n'a pas de sens en pratique (les mesures seraient inversées). C'est pour cela que d'autres restrictions ont été mises en place pour des raisons de bon sens. La première est la positivité de la pente, puisque l'exemple précédent crée des situations non pratiques. La deuxième restriction est une ordonnée à l'origine qui n'est pas trop négative afin de ne pas complètement « aplatiser » les valeurs. Par exemple : prenons un jeu de mesures qui s'échelonnent entre 10 et 1000 avec un seuil de positivité à 50. Si l'ordonnée à l'origine vaut -3, et en gardant une pente de 1, les nouvelles mesures s'échelonneront entre 0.1 et 10 et seront toutes sous le seuil de 50. C'est également un cas qui ne se retrouve pas en pratique.

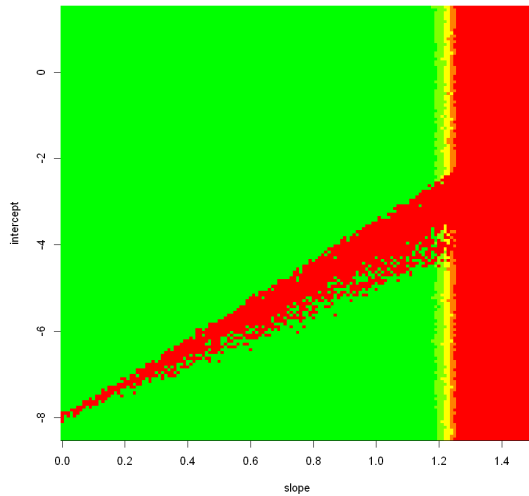


FIG. 4.1. : Tests d'équivalence

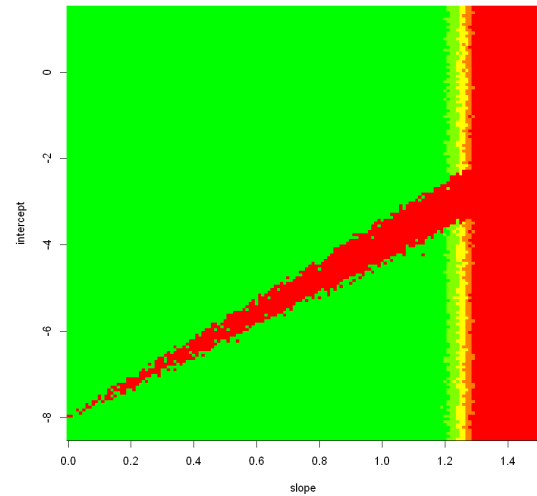


FIG. 4.2. : Tests de non-infériorité

## 4.5. Exemples

Le premier cas sur lequel j'ai travaillé est une étude randomisée ouverte (open-label randomized trial) portant sur l'hépatite A et B, et plus particulièrement sur la comparaison entre deux vaccins dans des groupes d'enfants de 12 à 15 ans. Le but de l'étude est de démontrer qu'un vaccin qui ne requiert que deux injections est au moins aussi efficace qu'un vaccin qui nécessite 3 injections. Les données d'intérêt, qui serviront de base pour les simulations, sont les titres mesurés après la dernière injection. Les deux critères utilisés pour l'étude sont le GMR et le SCR.

En particulier, pour chaque simulation réalisée, on calcule les intervalles de confiance d'un test de Student sur les GMT ainsi que ceux d'une différence de proportions pour les SCR. Les deux graphiques suivants représentent les résultats obtenus (à gauche en effectuant des tests d'équivalence et à droite en effectuant des tests de non-infériorité) avec le code de couleur suivant :

- zone verte : zone où 90% au moins des jeux de données simulés ne rejettent pas les hypothèses de tests (équivalence ou non-infériorité)
- zone intermédiaire : zone où entre 60 et 90% (par pas de 10%) des jeux de données simulés ne rejettent pas les hypothèses de tests (équivalence ou non-infériorité)
- zone rouge : zone où moins de 60% des jeux de données simulés ne rejettent pas les hypothèses de tests (équivalence ou non-infériorité)

Ces deux graphiques sont à peu près similaires et peuvent s'interpréter de la même manière, à savoir que l'on peut trouver une limite supérieure à la pente (entre 1.2 et 1.3 ici) et qui est en accord avec la spécification analytique précédente (la limite inférieure étant fixée à 0 au minimum). On observe également une zone rouge diagonale et dont l'explication pourrait tenir à la particularité du jeu de données de départ et au fait que pour les valeurs de l'ordonnée à l'origine concernées (plus petites que -2), les mesures sont écrasées et de petite particularité peuvent alors influencer le résultat final.

Le second cas traité est une grippe, et plus particulièrement sur la comparaison de deux vaccins dans des groupes d'enfants de 3 à 17 ans<sup>[4]</sup>. Le vaccin saisonnier contre la grippe est composé de deux virus de souches A et d'un virus de souche B, mais la souche B choisie (il en existe plusieurs et elle est prédite à partir des données précédentes) pour une saison n'est pas forcément celle qui sera présente dans le virus. Le but de l'étude est de démontrer qu'un vaccin quadrivalent avec deux souches A et deux souches B est au moins aussi efficace qu'un vaccin trivalent avec deux souches A et une souche B. Les données d'intérêt, qui serviront de base pour les simulations, sont les titres mesurés avant et après l'injection. Comme pour l'étude précédente, les deux critères utilisés pour l'étude sont le GMR et le SCR.

Puisque l'on dispose de données pré-vaccination, il faut en tenir compte lors de la comparaison des GMT et cela est fait à l'aide d'une ANCOVA suivie du calcul de l'intervalle de confiance d'un test post-hoc de Tukey sur les moyennes ajustées. Comme auparavant, on calcule également l'intervalle de confiance d'une différence de proportions pour les SCR. De même que pour l'étude précédente, les deux graphiques suivants représentent les résultats obtenus en gardant le même code de couleurs.

Contrairement à l'étude précédente, les deux graphiques présentés ci-dessus ne sont pas semblables. En effet, si le premier graphique (tests d'équivalence) permet de retrouver une limite supérieure pour la pente, ce n'est pas le cas pour le second où toutes les pentes positives sont acceptées. L'impact de l'ordonnée à l'origine est le même sur les deux graphiques, i.e. il n'y en a pas. Les échelles ont été volontairement modifiées afin d'insister sur le fait que même avec des valeurs relativement grandes, il n'y a pas de changement.

## 4.6. Conclusion

Les résultats des deux études sont similaires dans le sens où les conclusions tirées sont les mêmes, à savoir qu'il ne semble pas possible de déterminer des limites intéressantes pour l'ordonnée à l'origine et la pente. En effet, pour les deux études, l'ordonnée à l'origine ne semble pas modifier la conclusion du bridging alors que la pente semble être majorée par une valeur proche ( $\approx 1,3$ ) de celle calculée lors du développement analytique.

En particulier, la dissociation du graphe Total en deux graphes GMR et SCR/SPR

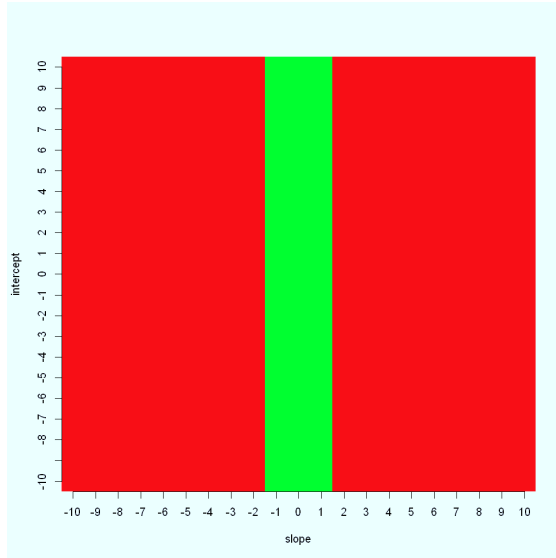


FIG. 4.3. : Tests d'équivalence

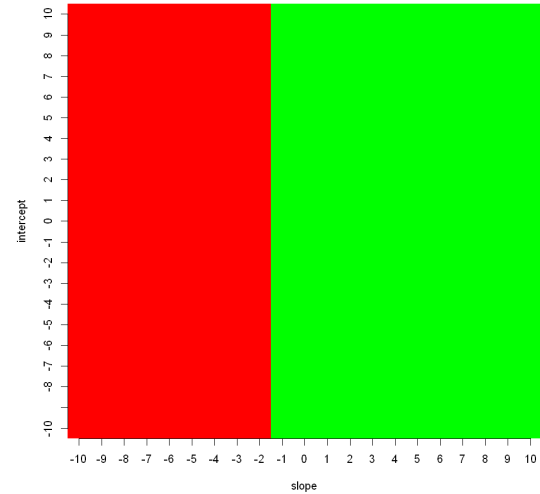


FIG. 4.4. : Tests de non-infériorité

indique que la borne supérieure d'acceptation de la pente provient du GMR et que le SPR ou SCR ne permet pas de déterminer une limite intéressante/pertinente ? pour l'ordonnée à l'origine.

Par ailleurs, et comme prévu par le développement analytique, la pente n'est pas bornée dans certains cas, notamment lors de la non-infériorité pour la seconde étude.

## 5. Bridging : partie technique

### 5.1. Présentation

La partie technique du bridging est consacrée aux choix des outils utilisés pour le bridging. Il est pour cela important de comprendre la problématique de la partie technique, à savoir qu'en terme de design, nous sommes face à un problème d'équivalence. Il s'agit en effet de montrer que deux manières de mesurer ne diffèrent pas sur le critère de mesure principal. Pour cela, plusieurs outils sont à notre disposition, et certains sont tout particulièrement adaptés à la comparaison de deux méthodes de mesures. Certains modèles de régression (appelés *errors-in-variables models* ou *measurement errors models*) permettent de tenir compte des erreurs sur la variable expliquée mais également sur la variable explicative, et c'est le point qui nous intéresse puisqu'il y a des erreurs de mesure sur les deux variables. Ces régressions seront présentées au cours des sections suivantes.

Le bridging technique peut être séparé en deux parties :

- L'étape qualitative : la proportion de mesures concordantes (les sujets testés positifs (respectivement négatifs) avec l'ancienne méthode qui restent positifs (respectivement négatifs) avec la nouvelle) du tableau de contingence doit être supérieure à un certain seuil (80% général). Par ailleurs, un test de McNemar est également réalisé sur ce tableau de contingence. Ce test permet de vérifier la symétrie du tableau par rapport à sa diagonale principale.
- L'étape quantitative : le GMR entre les mesures de l'ancienne et de la nouvelle méthode est calculé et doit être compris dans un certain intervalle prédéfini. Une régression de Deming (voir section suivante) est réalisée et l'intervalle de confiance de la pente doit également être compris dans un certain intervalle prédéfini. Enfin, on calcule des intervalles de prédiction, qui doivent se situer dans des intervalles d'acceptance prédéfinis. Plusieurs méthodes existent pour calculer ces intervalles de prédiction et c'est l'objet de la seconde partie de mon stage.

Si une nouvelle méthode de calcul des intervalles de prédiction a été choisie durant mon stage (par un sous-traitant), j'ai eu l'occasion d'effectuer un travail de recherche et de synthèse afin de présenter mes recommandations vis-à-vis de la méthode retenue.

## 5.2. Régression et modèles errors-in-variables

La méthode des moindres carrés ordinaires (OLS) est probablement la méthode d'estimation la plus connue que l'on applique à un modèle linéaire simple mais ses hypothèses, en particulier l'homoscédasticité et l'absence d'erreurs de mesure, ne sont pas toujours vérifiées. Cela peut conduire à une mauvaise estimation des paramètres et produire des résultats biaisés, ce qui arrive souvent lorsqu'il s'agit de comparer deux méthodes (cas du bridging). Une méthode améliorée de celle des moindres carrés est la méthode des moindres carrés pondérés (WLS) qui permet de tenir compte de l'hétéroscédasticité de la variable expliquée mais considère toujours que la variable explicative est sans erreur.

Les modèles de régression qui suivent permettent de s'affranchir de ces deux hypothèses que sont l'homoscédasticité des erreurs et l'exactitude de la variable explicative. Le modèle à rapport de variance constant (CVR) considère l'erreur de mesure pour les deux variables à l'aide du rapport des variances de la réponse et du prédicteur, supposé constant pour toutes les mesures, et appelé  $\lambda_{XY} = \sigma_Y^2 / \sigma_X^2$  où X et Y correspondent respectivement au prédicteur et à la réponse. En ce sens, ce modèle ne permet pas de tenir compte des erreurs de mesure individuelles. En dimension 2, le modèle CVR est aussi appelé régression de Deming (DR). Un cas particulier du modèle CVR est la régression orthogonale, également appelée modèle des moindres carrés totaux (TLS), où l'on considère que  $\lambda = 1$ , i.e. que les erreurs de mesures sont de même amplitude. L'idée de ce modèle est minimiser la distance euclidienne des points à la droite d'ajustement, là où la méthode OLS cherche à minimiser la distance verticale. Enfin, et c'est la méthode la moins contraignante de toutes puisqu'elle tient compte des erreurs de mesure individuelles ( $\lambda_{XY}$  n'est pas supposé constant), la régression BLS (bivariate least squares) semble être la meilleur choix pour comparer deux méthodes de mesure<sup>[3]</sup>.

Il faut savoir qu'en général,  $\lambda_{XY}$  n'est pas connu et qu'on estime  $\sigma_X^2$  et  $\sigma_Y^2$  par  $s_X^2$  et  $s_Y^2$  en utilisant des répliques<sup>[1]</sup>.

## 5.3. Intervalles de prédiction

En pratique, lorsqu'un bridging est réalisé, on considère pour l'instant que le rapport  $\lambda_{XY}$  est constant. Dans ce cas, on peut montrer que les estimateurs de  $\alpha$  et  $\beta$  (paramètres de la pente de régression) sont les mêmes pour les modèles CVR et BLS mais ces derniers se différencient lors des calculs d'intervalles de prédiction, et plus particulièrement sur l'estimation de la variance de la variable réponse et prédictive. C'est sur ce point que je me suis attardé puisqu'au cours de mon stage, la méthode de calcul de cet intervalle a été remise en question et confrontée à d'autres.

Il est à noter que le calcul des intervalles de prédiction requiert de calculer

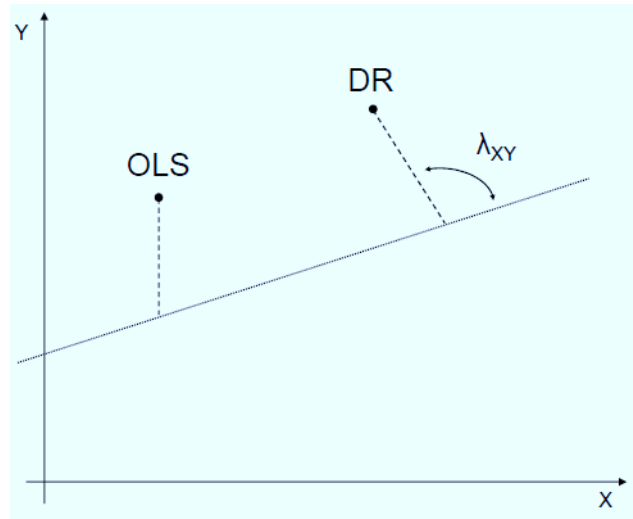


FIG. 5.1. : Différence de projection entre la méthode OLS et la régression de Deming

la variance des deux estimateurs  $\hat{\alpha}$  et  $\hat{\beta}$ .

Voici les différentes conclusions auxquelles je suis arrivé :

- Lorsque  $\lambda_{XY}$  est constant, il existe une méthode pour calculer la valeur exacte de l'intervalle de confiance de  $\beta$  mais elle n'est pas toujours utilisée car il peut arriver des situations impossibles où l'on cherche à calculer l'Arcsinus d'une valeur strictement plus grande que 1. Néanmoins, on peut utiliser la méthode de « plug in » pour insérer cette méthode dans une autre, notamment la méthode BLS<sup>[6]</sup>.
- Lorsque  $\lambda_{XY}$  est quelconque (constant ou non), le modèle BLS est celui qui fournit les meilleures prédictions en moyenne pour les variances de la variable réponse et prédicteur (4% et 5% contre 444% et 637% pour le modèle CVR, contre 1697% et 462% pour le modèle OR)<sup>[3]</sup>.

La méthode BLS est donc la méthode de calcul la plus adaptée et c'est celle-ci (en utilisant le calcul exact de l'intervalle de confiance de  $\beta$  à l'aide du « plug in ») qui a été retenue et est encore en cours d'implémentation.



## 6. Conclusion

Ce stage a été pour moi très enrichissant, tant sur un plan professionnel que social. J'ai eu la chance de travailler au sein de deux équipes différentes, et surtout dans une atmosphère très agréable tout au long. Les missions des deux équipes n'étant pas les mêmes, j'ai pu acquérir un certain nombre de connaissances sur le domaine qu'est la vaccinologie mais aussi sur l'organisation de l'équipe, qui doit être adaptée aux objectifs ciblés (une phase exploratoire n'est pas une phase de développement). Il n'est pas sans préciser que ce travail a permis de développer mes compétences en statistiques.

Durant ce stage, j'ai pu découvrir les enjeux et les attentes de la recherche en vaccinologie ainsi que les besoins d'utiliser des méthodes statistiques répondant à des concepts biologiques. J'ai pu également participer aux différentes étapes d'un projet, partant de la phase de recherche bibliographique jusqu'à l'implémentation des méthodes.

Par ailleurs, le travail de recherche sur une nouvelle approche de bridging m'a permis de découvrir un domaine d'application des statistiques que je ne connaissais pas, et qui ne possède pas une méthodologie universelle, à savoir la comparaison de méthodes. L'enjeu du bridging se situe à la fois sur un plan économique, chronologique mais aussi biologique et la recherche d'une méthode qui permettrait d'économiser des ressources sur ces trois plans est un projet prioritaire de toute évidence.

J'ai par ailleurs découvert la manière dont se déroule la gestion d'un projet avec toutes ses contraintes administratives et organisationnelles. Contraintes d'autant plus accentuées que GSK Vaccines est une des plus grandes entreprises dans son domaine et possède une hiérarchie en conséquence. Walthère a été un maître de stage exemplaire, puisqu'il a non seulement favorisé mon introduction aux différentes équipes, mais a toujours trouvé du temps lorsque j'avais besoin de son aide, notamment pour la compréhension du contexte en début de stage. J'ai donc pu approfondir et appliquer certaines de mes connaissances acquises lors de ma formation et pratiquer les langages R et SAS.

Ces six mois de stage ont été pour moi l'opportunité de découvrir le métier de biostatisticien en situation réelle, au sein de l'entreprise GSK Vaccines, et demeurera une excellente expérience.

## Bibliographie

- [1] Bernadette B. Gobaerts Bernard G. Francq. Hyperbolic confidence bands of errors-in-variables regression lines applied to method comparison studies. *Journal de la Société Française de Statistique*, 155(1) :23–45, 2014.
- [2] Patrick Gagnon. Intervalles de confiance pour une différence de deux proportions. Master's thesis, Faculté des études supérieures et postdoctorales, Université Laval, 2006.
- [3] Xavier Rius Javier del Río, Jordi Riu. Prediction intervals in linear regression taking into account errors on both axes. *Journal of Chemometrics*, 15 :773–788, 2001.
- [4] Joanne M. Langley et al. Immunogenicity and safety of an inactivated quadrivalent influenza vaccine candidate : A phase III randomized controlled trial in children. *The Journal of Infectious Diseases*, 208(4) :544–553, 2013.
- [5] Jozef Nauta. *Statistics in Clinical Vaccine Trials*. Springer, 2011.
- [6] Charles Y. Tan and Boris Iglewicz. Measurement-methods comparisons and linear statistical relationship. In *Technometrics*, volume 41, pages 192–201. 1999.
- [7] Xiao-Feng Wang and Bin Wang. Deconvolution estimation in measurement error models : The r package decon. *Journal of Statistical Software*, 39(10), 2011.
- [8] Xiao-Feng Wang and Bin Wang. *Deconvolution Estimation in Measurement Error Models*, 2013.

## A. Simulations

On note  $X_{i,old}, i \in [1; n]$  les anciennes mesures des titres et  $X_i, i \in [1; n]$  les nouvelles mesures simulées où  $n$  représente le nombre total (pré + post pour chaque groupe) de mesures effectuées et gardées. En notant  $Y_i, i \in [1; n]$  les mesures simulées avec un intercept  $\alpha$  et une pente  $\beta$  par l'hypothèse de linéarité, on utilise les relations suivantes dans la simulation « par perturbation » :

$$\log_{10}(X_i) = \log_{10}(X_{i,old}) + \epsilon, \epsilon \sim \mathcal{N}(0, 0.0860)$$

$$\log_{10}(Y_i) = \alpha + \beta \cdot \log_{10}(X_i) + \epsilon, \epsilon \sim \mathcal{N}(0, 0.0860)$$

La valeur de l'écart-type, à savoir 0.0860, permet de simuler les erreurs de mesure et tient compte de la log-transformation des valeurs. Plus précisément, cette valeur est déduite à partir de la formule suivante (valable pour une variable aléatoire suivant une loi log-normale de variance  $\sigma^2$ ) :

$$CV = \sqrt{e^{\sigma^2} - 1}$$

En effet, si  $U$  est une variable aléatoire de loi log-normale de paramètres  $(\mu, \sigma)$ , alors sa densité est  $\forall x > 0, f_U(u) = \frac{1}{u \cdot \sigma \cdot \sqrt{2 \cdot \pi}} \exp(-\frac{\ln(u-\mu)}{2 \cdot \sigma^2})$ , sa moyenne  $E(U) = \exp(\mu + \frac{\sigma^2}{2})$ , et son écart-type  $Sd(U) = \exp(\mu + \frac{\sigma^2}{2}) \cdot \sqrt{\exp(\sigma^2) - 1}$ , alors en notant  $CV(U)$  le coefficient de variation de  $U$ , on a :

$$\begin{aligned} CV(U) &= Sd(U)/E(U) \\ &= \sqrt{\exp(\sigma^2) - 1} \\ \Leftrightarrow \sigma^2 &= \ln(1 + CV^2(U)) \end{aligned}$$

## B. Approche analytique

Notations utilisées dans cette partie :

- $X$  : anciennes mesures
- $Y$  : nouvelles mesures
- $X_k$  : anciennes mesures du groupe k
- $Y_k$  : nouvelles mesures du groupe k
- $U_{i,k}$  : i-ième individu du k-ième groupe des mesures U
- $n_k$  : nombre de sujets dans le groupe k
- $GMT_{U,k}$  : GMT du groupe k des mesures U
- $IC$  : intervalle de confiance de  $GMR_X$
- $Lower(GMR_X)$  : borne inférieure de l'intervalle de confiance de  $GMR_X$  (noté  $L_X$  par soucis de clareté parfois)
- $Upper(GMR_X)$  : borne supérieure de l'intervalle de confiance de  $GMR_X$  (noté  $U_X$  par soucis de clareté parfois)
- $[\lambda_L, \lambda_U]$  : intervalle fixé arbitrairement pour les tests d'équivalence sur les GMT (par exemple, il faut que l'intervalle de confiance du GMR soit dans cet intervalle pour accepter l'équivalence)
- $L = \frac{\log(\lambda_L)}{\log(L_{GMR_X})}$
- $U = \frac{\log(\lambda_U)}{\log(U_{GMR_X})}$

On a,

$$\begin{aligned}
 GMT_Y &= 10^{(\frac{1}{n} \sum_{i=1}^n (\log(Y_i)))} \\
 &= 10^{(\frac{1}{n} \sum_{i=1}^n (\alpha + \beta \cdot \log(X_i)))} \\
 &= 10^\alpha \cdot 10^{(\beta \cdot \frac{1}{n} \sum_{i=1}^n (\log(X_i)))} \\
 &= \alpha' \cdot GMT_X^\beta
 \end{aligned}$$

$$\text{d'où } GMR_Y = \frac{GMT_{Y,T}}{GMT_{Y,C}} = \frac{\alpha' GMT_{X,T}^\beta}{\alpha' GMT_{X,C}^\beta} = GMR_X^\beta$$

A partir de cette relation, on peut traiter les trois cas (non-infériorité, équivalence, supériorité) séparément en remarquant que  $Lower(GMR_X^\beta) = Lower^\beta(GMR_X)$ . En effet :

$$\begin{aligned} GMT(X_1) &= 10^{(\frac{1}{n_1} \sum_{i=1}^{n_1} (\log(X_{i,1})))} \\ GMT(X_2) &= 10^{(\frac{1}{n_2} \sum_{i=1}^{n_2} (\log(X_{i,2})))} \\ GMR(X) &= 10^{(\frac{1}{n_1} \sum_{i=1}^{n_1} (\log(X_{i,1})) - \frac{1}{n_2} \sum_{i=1}^{n_2} (\log(X_{i,2})))} \end{aligned}$$

d'où

$$\begin{aligned} (\overline{X_1} - \overline{X_2}) \pm t_{\eta/2} \cdot Sd(\overline{X_1} - \overline{X_2}) &= [\log(Lower(GMR_X)), \log(Upper(GMR_X))] \\ (\overline{Y_1} - \overline{Y_2}) \pm t_{\eta/2} \cdot Sd(\overline{Y_1} - \overline{Y_2}) &= (\overline{X_1} - \overline{X_2}) \pm t_{\eta/2} \cdot Sd(\overline{X_1} - \overline{X_2}) \\ &= \beta \times [\log(Lower(GMR_X)), \log(Upper(GMR_X))] \end{aligned}$$

$$\text{D'où } 10^{Lower(GMR_X)} \leq GMR_X \leq 10^{Upper(GMR_X)} \text{ et } 10^{\beta \cdot Lower(GMR_X)} \leq GMR_Y \leq 10^{\beta \cdot Upper(GMR_X)} \Leftrightarrow (10^{Lower(GMR_X)})^\beta \leq GMR_X^\beta \leq (10^{Upper(GMR_X)})^\beta$$

Puis  $Lower(GMR_X) \leq GMR_X \leq Upper(GMR_X)$  et  $Lower^\beta(GMR_X) \leq GMR_Y \leq Upper^\beta(GMR_X) \Leftrightarrow Lower^\beta(GMR_X) \leq GMR_X^\beta \leq Upper^\beta(GMR_X)$ , ce qui démontre le résultat recherché. On peut dès lors intégrer ce résultat dans les trois cas suivants où la première inégalité correspond à un test de non-infériorité, la deuxième (combinée avec la première) à un test d'équivalence et la dernière à un test de supériorité :

$$\begin{aligned} \lambda_L &\leq L_Y \\ \Leftrightarrow \lambda_L &\leq L_X^\beta \\ \Leftrightarrow \log(\lambda_L) &\leq \beta \cdot \log(L_X) \\ \Leftrightarrow \frac{\log(\lambda_L)}{\log(L_X)} &\leq \beta \end{aligned}$$

$$\begin{aligned} U_Y &\leq \lambda_U \\ \Leftrightarrow U_X^\beta &\leq \lambda_U \\ \Leftrightarrow \beta \cdot \log(U_X) &\leq \log(\lambda_U) \\ \Leftrightarrow \beta &\leq \frac{\log(\lambda_U)}{\log(U_X)} \end{aligned}$$

$$\begin{aligned} 1 &\leq L_Y \\ \Leftrightarrow 1 &\leq L_X^\beta \\ \Leftrightarrow 0 &\leq \beta \cdot \log(L_X) \\ \Leftrightarrow \frac{0}{\log(L_U)} &\leq \beta \end{aligned}$$

Note : le cas pratique ci-dessous décrit essentiellement la position de l'intervalle de confiance de  $GMR_X$  par rapport à 0 dans le cas où l'hypothèse testée est vérifiée.

Tab. B.1. : Spécifications sur  $\beta$  selon les cas

Cas pratique	Equivalence	Non-Infériorité	Supériorité
$0 \in IC$	$\beta < \min(L, U)$	$\beta < L$	
$IC < 0$	$\beta \in [U, L]$	$\beta < L$	
$IC > 0$	$\beta \in [L, U]$	$\beta > L$	$\beta > 0$
$Lower > 0$			$\beta > 0$