



HAL
open science

Approche statistique tenant compte de la variabilité analytique dans l'évolution de la cinétique d'une réponse immunitaire

Jérémy Magnanensi

► To cite this version:

Jérémy Magnanensi. Approche statistique tenant compte de la variabilité analytique dans l'évolution de la cinétique d'une réponse immunitaire. *Méthodologie [stat.ME]*. 2010. dumas-00520265

HAL Id: dumas-00520265

<https://dumas.ccsd.cnrs.fr/dumas-00520265v1>

Submitted on 22 Sep 2010

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 Biologicals
Influenza Vaccines Department
Avenue Fleming, 20
B-1300 Wavre, Belgium

Université de Strasbourg
U.F.R. de Mathématique
7, rue René Descartes
67084 Strasbourg

Rapport de Stage
Master de Statistique 1ère Année
31 Mai 2010 - 20 Août 2010

Intitulé du stage : Approche statistique tenant compte de la variabilité analytique dans l'évolution de la cinétique d'une réponse immunitaire.

Stagiaire:

MAGNANENSI Jérémy
Stagiaire bio-statisticien

Maître de Stage :

DEWE Walthere P.
Senior Manager Bio-Statistician – Clinical Biometrics

Table des Matières :

Remerciements

I. Introduction

II. Présentation de l'entreprise

2.1. Glaxo Smith Kline (GSK), un leader mondial

2.2. GSK Biologicals en Belgique

III. Contexte et problème posé

3.1. Contexte et explication de l'étude

3.1.1. Origine du problème

3.1.2. Planification de l'étude utilisée

3.2. Objectif et informations générales

3.2.1. Objectif

3.2.2. Hypothèse de travail

3.2.3. Logiciel

3.2.4. Confidentialité

3.2.5. Modèles statistiques utilisés

3.2.6. Informations préliminaires

IV. Développement et solutions apportées

4.1. Approche fréquentiste

4.1.1. Idée générale de l'approche

4.1.2. Corrections des titres

4.1.3. Estimation des variances et intervalles de confiance

4.1.4. Résultats

4.2. Approche bayésienne

4.2.1. Motivation

4.2.2. Inconvénient

V. Conclusion

5.1. Questionnements au sujet des résultats

5.2. Bilan général du stage

Annexes

Remerciements :

En premier lieu, je tenais à remercier Mr Dewe Walthere, mon maître de stage, pour son écoute, ses aides et la confiance qu'il m'a porté dans l'exécution quotidienne de mes tâches tout au long du stage. Il a pris tout particulièrement soin à ce que mon intégration au sein du service de statistique Flu se passe au mieux et a toujours été disponible en cas de problèmes lors de la résolution de mes objectifs.

Ensuite, je tenais également à remercier tout particulièrement Mlle Dupuy Laure, ainsi que nos colocataires Anouk et Raf, pour leur accueil et leur hospitalité tout au long de ces trois mois passés à Bruxelles.

Je remercie notamment Mr Tibaldi Fabian, Senior Manager in Clinical Biometrics, pour l'aide et les conseils qu'il nous a apporté sur ce sujet ainsi que Mr Fourneau Marc, Directeur GRCD Biometrics pour la confiance qu'il m'a porté en validant mon stage au sein de GSK.

Enfin, je tiens naturellement à remercier ma famille, mon père, ma mère, ma soeur et mon frère pour leur soutien qu'ils m'ont apporté afin que je puisse réaliser mon stage dans les meilleures conditions possibles.

I. Introduction

Lors de mon année scolaire de Master 1, j'ai décidé d'axer mes recherches de stage dans le domaine de la bio-statistique. J'ai donc recherché un stage et envoyé des lettres aux principaux laboratoires pharmaceutiques ainsi qu'aux C.R.O. (Contract Research Organisation), entreprises spécialisées dans les études cliniques pour le compte de l'industrie pharmaceutique. J'ai ainsi eu un retour positif de l'entreprise pharmaceutique GlaxoSmithKline basée à Wavre en Belgique, en périphérie de Bruxelles. C'est sans aucune hésitation que j'ai accepté cette offre de stage, me donnant ainsi l'opportunité de me faire ma propre idée sur ce secteur et comprendre comment cela se passe au sein d'une entreprise pharmaceutique, qui plus est, chez le numéro deux mondial.

Ce stage m'a d'autant plus intéressé qu'il a mêlé recherche et application direct. De plus, cela m'a permis d'effectuer cette période de ma vie universitaire à l'étranger, ce qui humainement parlant est loin d'être négligeable. Bruxelles est une ville formidable, avec beaucoup de vie, d'animations culturelles, de mixité sociale et les habitants y sont très chaleureux. En ce qui concerne mon poste chez GSK, j'ai travaillé au sein d'une équipe statistique, sur un projet lié à l'analyse de la variabilité inter-run de certains essais cliniques. En effet, lorsque des échantillons de sang sont prélevés chez des sujets d'étude clinique, ils peuvent être analysés à des moments différents, c'est-à-dire dans des séries, run en anglais, différentes. La répartition de l'analyse en série implique une source de variabilité supplémentaire, l'effet run. Dans le cas où certaines analyses sont répétées d'une série à l'autre, l'objectif travail était d'évaluer la possibilité de réduire l'impact de cette source de variation sur l'estimation de la cinétique de la réponse immunitaire. Il s'agissait donc d'isoler l'effet inter-run puis de corriger les titres de façon correcte afin de pouvoir procéder à une analyse statistique correct des données et pouvoir ainsi donner une conclusion sur l'effet du vaccin testé sans que celle-ci soit faussée par une variabilité d'un effet qui n'est pas en lien avec le médicament testé. Ceci fut très intéressant dans le sens où il ne s'agissait pas uniquement de faire l'analyse de données, il y avait une réelle part de recherche puisqu'il s'agissait d'un sujet non encore étudié. Ce stage, j'en étais persuadé, allait m'apporter beaucoup de choses, autant sur l'aspect du fonctionnement d'une entreprise, que sur les méthodes de travail utilisées dans le milieu professionnel et le moins que je puisse dire c'est que je ne fus pas déçu. Il y'a une réelle différence dans les méthodes de travail entre le milieu universitaire ou scolaire et le milieu professionnel.

Pour ma part, j'ai beaucoup appris sur le fonctionnement de laboratoires pharmaceutiques, non pas que sur mon sujet de stage, mais sur divers autres sujets liés à l'entreprise tel que l'organisation d'une étude, le suivi d'une étude clinique, l'analyse statistique des résultats, comment se préparent les tests, le plan d'expérience, sans pour autant en avoir fait mais simplement en parlant de cela avec les professionnels sur place. GSK est une entreprise très dynamique, composée d'un personnel, du moins au sein du service statistique dans lequel j'ai effectué ce stage, très ouvert et rigoureux dans leur travail. Cette entreprise a donc été pour moi un très bon cadre de travail, idéal pour y effectuer ce stage. De plus, je désirais vraiment effectuer mon stage dans une entreprise où je pourrai côtoyer des personnes spécialistes des statistiques et non une entreprise où l'on aurait compté que sur moi pour effectuer ma mission et résoudre un problème sans pouvoir apprendre d'une personne spécialiste du domaine. Tout ceci je l'ai trouvé chez GlaxoSmithKline, avec des personnes très accessibles et qui ont su prendre le temps de m'aider ou de m'expliquer certains points lorsque j'avais des difficultés.

II. Présentation de l'entreprise

2.1. Glaxo Smith Kline (GSK), un leader mondial

GSK est l'un des leaders mondiaux dans le domaine pharmaceutique. En effet, il s'agit du deuxième laboratoire pharmaceutique mondial, en terme de chiffre d'affaire, en affichant une forte présence dans une vingtaine de domaines thérapeutiques tel que l'oncologie, la neurologie, le VIH sans oublier la vaccinologie dont les vaccins antigrippaux, secteur dans lequel j'ai effectué mon stage.

Il s'agit d'une entreprise possédant un très grand passé. En effet, son historique remonte à 1715 avec la création, à Londres, de la pharmacie de Plough Court par Silvanus Bevan. En 1830, John K. Smith ouvrira sa première pharmacie à Philadelphie qui, rejoint par son frère en 1841, formeront John K Smith & Co. S'en suivront de multiples fusions et associations qui feront naître en 2000 GSK à travers la fusion de Glaxo Wellcome et SmithKline Beecham.

GSK emploie aujourd'hui quelques 100.000 personnes à travers le monde en étant présent dans près de 120 pays. Elle est constituée de deux parties à savoir GSK Santé Grand Public, spécialisée dans les médicaments d'automédication vendus sans prescription en pharmacie et non remboursables ainsi que dans le domaine buccodentaire, et GSK Pharma, spécialisée dans les médicaments vendus sur prescription médicale, et à l'intérieur de laquelle se situe GSK Biologicals focalisée sur les vaccins. C'est donc au sein de GSK Bio que j'ai effectué mon stage, plus particulièrement dans le secteur s'occupant des virus anti-grippaux. GSK Bio c'est plus de 9000 personnes dans le monde dont 1600 scientifiques qui développent des vaccins pour le monde entier et notamment des vaccins contre les trois principaux fléaux planétaires, considérés comme prioritaires par l'OMS, à savoir le VIH/SIDA, la tuberculose et enfin le paludisme. GSK Biologicals abrite trois sites de production belges, le premier et plus ancien à Rixensart, puis un site à Gembloux et un dernier à Wavre, site sur lequel s'est déroulé mon stage. Ces trois sites réunis constituent l'un des plus importants pôles industriels de vaccins en Europe et accueillent également l'ensemble des services qui coordonnent l'activité de recherche, de développement et de production de GSK Biologicals dans le monde. GSK Biologicals, c'est aussi 14 sites de production sur 3 continents, qui chaque année produisent, formulent, conditionnent, contrôlent et expédient plus d'un milliard de doses d'une trentaine de vaccins différents dans 169 pays en sachant que près de 80% de la production est destinée aux pays en développement. Ces vaccins permettent d'éviter plus de trois millions de décès chaque année et des séquelles liées à des maladies infectieuses à plus de 750.000 enfants. Enfin, GSK possède un chiffre d'affaire de 34.1 milliards d'euros et en consacre 17% au seul secteur de Recherche & Développement. GSK Pharma représente 29.5 milliards d'euros et GSK Santé Grand Public 4.6 milliards. A noter que 445 millions d'euros sont consacrés à des programmes humanitaires soutenus par GSK et sa fondation.

En effet, GSK est une entreprise très engagée dans le suivi des patients, notamment des enfants et des populations du tiers-monde. Grâce, entre autre, à sa fondation, GSK développe des actions dans plusieurs domaines. En effet, ils sont présents dans l'accompagnement des enfants malades, le développement de la recherche dans le domaine de la santé des Femmes, l'amélioration de l'accès à l'information pour les malades ou encore l'amélioration de la santé dans les pays en développement. Pour cela, ils se fixent des domaines de Recherche & Développement sur des maladies frappant majoritairement le tiers-monde, fournissent à ces pays les vaccins, antirétroviraux et antipaludéens

à tarifs préférentiels ou encore soutiennent des programmes de proximité concernant des maladies qui sévissent plus particulièrement dans ces pays comme la filariose lymphatique, le VIH/SIDA, le paludisme ainsi que la diarrhée de l'enfant en procédant à des dons en médicaments, financiers ou technique. Afin de faciliter la mise à disposition de ses vaccins, GSK a également étendu ses partenariats avec l'Unicef, la Global Alliance for Vaccine Immunization (GAVI), l'Organisation mondiale de la Santé (OMS) et l'Organisation Panaméricaine de la Santé (PAHO).

2.2. GSK Biologicals en Belgique

Tout d'abord, débutons par un petit historique. L'histoire commence par la création en 1945 de l'entreprise belge Recherche et Industrie Thérapeutiques (RIT) à Genval dans la banlieue de Bruxelles par le Docteur Pieter De Sommer. En 1956, débute la production du vaccin contre la polio qui va connaître un très grand succès d'où l'acquisition dès 1958 du site de Rixensart. En 1968, RIT est racheté par SmithKline & French et devient ainsi SmithKline-RIT avant de s'appeler en 1989 SmithKline Beecham Biologicals suite à la fusion de SmithKline et Beecham. En 1995, se produit l'extension aux deux sites supplémentaires de Wavre et Gembloux avant de finir par s'appeler GSK Biologicals suite à la dernière fusion en date de SmithKline et Glaxo en 2000.

Le site de Wavre est très récent et n'est même pas encore complètement achevé. Ce site emploie, au quotidien, aux alentours de 4000 personnes et est constitué de plus d'une quinzaine de bâtiments dont certains, comme celui dans lequel je travaillais, ont été inaugurés au courant des années 2009/2010. Il s'agit d'un site de production mais également de recherche regroupant beaucoup de domaines d'études ainsi que des services commerciales, financiers... En cumulant les 3 sites présents en Belgique, GSK Bio emploie, au mois de Septembre 2009, plus de 6800 personnes.

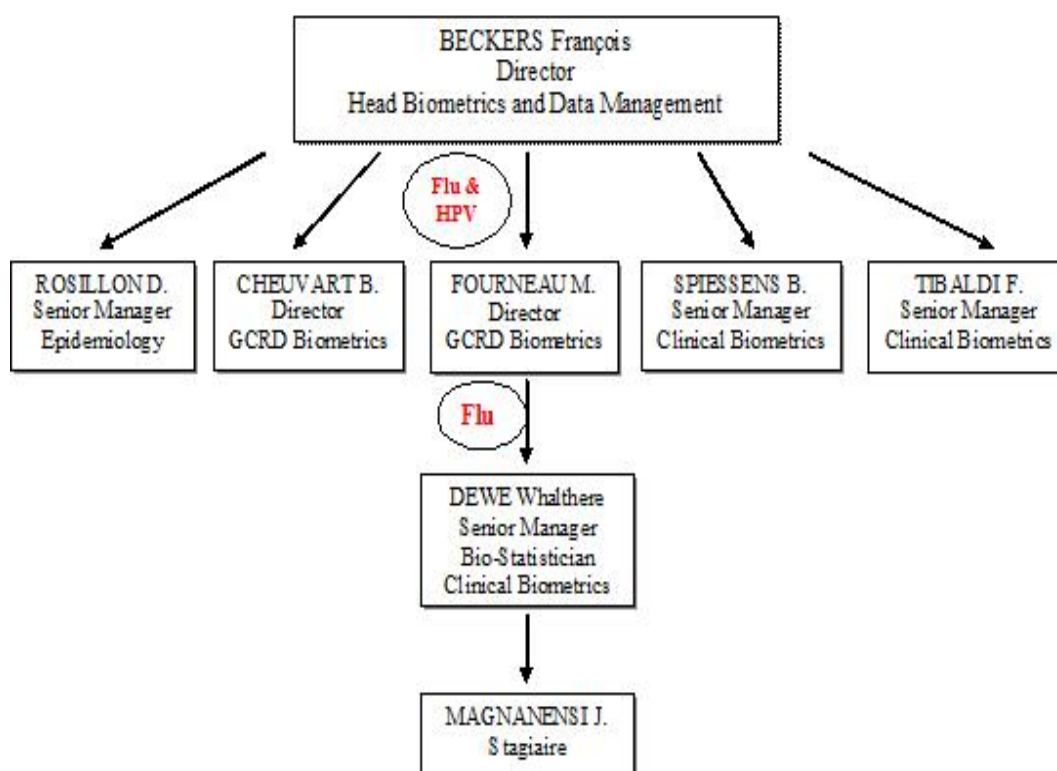
En ce qui concerne le service des statistiques plus précisément, il est constitué de quatre parties majeures :

- Epidémiologie
- Oncologie
- Pédiatrie
- Adult and Early Development

Le service statistique, regroupant ces quatre parties plus d'autres projets, est dirigé par Mr Beckers François.

Pour ma part, je travaillais au sein du groupe Flu (Grippe), groupe qui se trouve sous la direction de Mr Fourneau Marc et managé par mon maître de Stage, Mr Dewe Walthere. Ce groupe travail plus particulièrement sur l'analyse des données cliniques relatives aux vaccins contre les différentes gripes.

Diagramme de l'Organisation du Service Statistique :



III. Contexte et problème posé

3.1. Contexte et explication de l'étude

3.1.1. Origine du problème

Mon travail chez GSK était basé sur l'étude de l'impact de la variabilité inter-run sur l'évolution de la cinétique de la réponse immunitaire après vaccination. En effet, une difficulté apparaît si, lors du suivi des sujets, pour différentes raisons, on est amené à devoir faire des séries d'analyses intermédiaires à différents moments de l'étude, chaque série constituant un Run. Ceci est effectué dans le but de suivre l'évolution des paramètres étudiés à travers le temps, tout en obtenant le plus rapidement possible les premiers résultats. En effet, cette difficulté provient du fait que certains facteurs aléatoires et pas forcément contrôlables vont influencer les résultats des analyses suivant le moment où l'analyse sera pratiquée et pouvant ainsi rendre biaisée ou moins précise l'estimation de la cinétique. Le cas échant, il faut donc être capable de pouvoir comparer ces résultats de façon correcte c'est-à-dire en tentant d'éliminer ou du moins de tenir compte de cette variabilité inter-run.

3.1.2. Planification de l'étude utilisée

La résolution de ce problème d'estimation s'est basée tout d'abord sur une unique étude réalisée par GSK. Il s'agit d'une étude visant à mesurer l'évolution d'anticorps suite à l'administration d'un vaccin. Cette étude contient 130 patients, séparés en deux groupes, de respectivement 64 et 66 individus ayant chacun reçu un vaccin différent. La planification de l'étude fut la suivante :

Les individus devaient se présenter au Jour 0 sans avoir été déjà vaccinés au cours de l'année ni avoir été malades de la grippe au cours de l'année. A ce moment là, une prise de sang a été effectuée, l'échantillon conservé et congelé puis on leurs a administré une première dose de vaccin. Trois semaines plus tard, les individus sont revenus, on leurs refait une prise de sang au Jour 21 puis une deuxième dose leurs a été injectée. Ceci est réalisé afin de mesurer l'utilité d'une double vaccination sur la population. A ce moment là, le laboratoire effectue la première série d'analyses de sang (Run 1) qui renvoie pour chacun des individus deux titres, le premier indiquant la quantité d'anticorps au Jour 0 et le deuxième la quantité d'anticorps au Jour 21. Ces individus reviennent enfin au Jour 35 où une dernière prise de sang est réalisée. Ces prises de sang fournissent ainsi une troisième et dernière mesure pour chaque individu correspondant à la quantité d'anticorps au Jour 35, c'est à dire avec les deux injections de vaccin. Il s'agit là du Run 2.

Exemple d'une telle base de données.

Groupe	Sujet	Jour	Run	Titre
1	1	0	1	...
1	1	21	1	...
1	1	35	2	...
1	2	0	1	...
1	2	21	1	...
1	2	35	2	...
...
1	64	0	1	...
1	64	21	1	...
1	64	35	2	...
2	65	0	1	...
2	65	21	1	...
2	65	35	2	...
...
2	130	0	1	...
2	130	21	1	...
2	130	35	2	...

Pour cette étude, cette planification a induit le problème suivant. Lors du retour des résultats d'analyse, les titres d'un groupe avaient chuté au Jour 35 ce qui indiquait soit un problème lié au vaccin soit un problème lié aux conditions de l'analyse. Cependant un bridge a été effectué lors de l'analyse des échantillons de sang du Jour 35, c'est-à-dire un deuxième run au Jour 0 et au Jour 21 sur un échantillon d'individus choisis aléatoirement. Quarante mesures ont été réalisées avec ce bridge (on n'en retiendra que 29, les autres mesures se situant en-dessous d'un seuil d'acceptabilité), pouvant appartenir au Jour 0 ou au Jour 21 et au Groupe 1 ou au Groupe 2. L'étude des échantillons des jours 0 et 21 retestés dans le Run 2 indiquait la même diminution et laissait ainsi présager un problème d'analyse. Une troisième série d'analyse consistant à tout réanalyser confirmait cette hypothèse (Voir Annexe 1). Cependant, cette solution de réanalyse est coûteuse en temps, en argent et en énergie. La recherche d'un ajustement statistique prend donc tout son sens.

Exemple d'une base de données de bridging.

Groupe	Sujet	Jour	Run	Titre
1	24	21	1	...
1	24	21	2	...
2	78	21	1	...
2	78	21	2	...
...
1	17	0	1	...
1	17	0	2	...

La question qui se posait donc à moi était de savoir s'il serait possible de réduire le biais lié à l'effet run ou de tenir compte de sa variabilité analytique dans les résultats à l'aide des données du bridge?

3.2. Objectif et informations générales

3.2.1. Objectif

Le problème majeur de cette méthode est le temps qu'elle fait perdre au laboratoire d'analyse. En effet, celui-ci n'a pas le temps de recommencer, pour chaque étude, un Run 3 servant de confirmation.

L'objectif de mon travail était donc de trouver, à partir des données initiales ainsi que des données du bridge, une façon d'éliminer ou de réduire au maximum le biais potentiel lié à cet effet run ou de tenir compte de la variabilité analytique liée à cet effet. Ceci dans le but d'éviter, à l'avenir, de devoir procéder à un Run 3 de confirmation.

3.2.2. Hypothèse de travail

Il est supposé et admis que les titres HI (Hemagglutination Inhibition) suivent une loi log normale de base 10. Ainsi, mon travail a été basé sur l'étude des $\log_{10}(\text{HI})$ puisque suivant ainsi une loi Normale. Aucune vérification ni aucun test n'a été effectué par ma part sur ce sujet. En effet, il s'agit là d'une convention avec laquelle GSK travaille depuis des années et il n'aurait pas été raisonnable de changer cela. Il m'a donc été demandé de l'admettre.

3.2.3. Logiciels

Le seul logiciel que j'ai utilisé afin de réaliser mon stage, est le logiciel de statistique et de gestion de bases de données SAS version 9.2. Les procédures les plus fréquemment utilisées furent :

```
PROC MIXED  
PROC NLIN  
PROC GENMOD  
PROC GLM  
PROC MEANS  
PROC UNIVARIATE
```

J'ai également eu l'occasion de créer des macros et de faire de la gestion de tableaux ainsi que de la création et modélisation graphique.

Cependant, certains graphiques que j'ai produits ont été réalisés à l'aide du logiciel de gestion de données et tableur Microsoft Office Excel 2007. Ce rapport de stage a quant à lui été réalisé avec le logiciel L^AT_EX ainsi que Microsoft Office Word 2007.

3.2.4. Confidentialité

Toutes les données brutes et résultats d'analyses cliniques relatifs aux études sur lesquelles j'ai pu travailler sont confidentiels et n'apparaissent donc pas dans ce rapport de stage. Seuls les formules mathématiques théoriques et les résultats de ces formules utilisant ces données sont autorisées à être révélés dans ce rapport, tels que les effectifs, les moyennes, les variances...

3.2.5. Modèles statistiques utilisés

On est en face d'une étude où des mesures ont été collectées pour trois temps précis. Le premier temps d'étude est au jour 0, le deuxième au jour 21 et le troisième au jour 35. Dans ces données, on a enlevé les données du bridge que l'on traite à part. Ainsi, on va bien chercher un modèle pour les $\log_{10}(\text{HI})$ non corrigés et appartenant au Run1 pour le Jour0 et Jour21 et au Run2 pour le Jour35. L'effet jour est en fait mélangé à l'effet run. En effet, il y a une très forte corrélation entre ces deux variables avec une p-valeur significative au seuil de 5%, d'où le but d'isoler cet effet run à l'aide des données du bridge afin de pouvoir reprendre ensuite ce modèle pour les $\log_{10}(\text{HI})$ corrigés et estimer correctement tous les effets. (Voir Annexe 3 Fig.1 et 2)

On est ici face à un modèle à mesures répétées à deux facteurs fixes (le groupe et le jour) avec des répétitions sur les sujets suivant le facteur jour. Pour étudier le problème posé, on a donc choisi le modèle suivant:

$$(1) \quad Y_{ijs} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_s + \varepsilon_{ijs}, \quad \text{avec } 1 \leq i \leq 3, 1 \leq j \leq 2, \text{ et } 1 \leq s \leq 130.$$

où Y_{ijs} représente le $\log_{10}(\text{HI})$ avec les HI corrigés ou non-corrigés

α_i représente l'effet fixe du i-ème temps d'étude

(mélangé à l'effet run si la variable dépendante concerne les HI non-corrigés)

β_j représente l'effet fixe du j-ème groupe (\sim vaccin)

γ_s représente l'effet aléatoire du s-ème sujet

ε_{ijs} l'erreur résiduelle.

On impose ici comme contrainte que :

$$\alpha_3 = 0; \beta_2 = 0; (\alpha\beta)_{31} = (\alpha\beta)_{12} = (\alpha\beta)_{22} = (\alpha\beta)_{32} = 0; \gamma_s \sim N(0, \sigma_\gamma^2); \varepsilon_{ijs} \sim N(0, \sigma^2).$$

En ce qui concerne le bridge, on a des données qui ont été établies à partir d'échantillons de sang de 29 sujets choisis aléatoirement. Ces échantillons datent soit du jour 0 soit du jour 21 et appartiennent naturellement soit au groupe 1 soit au groupe 2. Il s'agit ainsi, comme expliqué auparavant, d'un tableau de données contenant pour chacun des 29 sujets deux mesures pour un jour précis, une première mesure qui n'est rien d'autre que celle obtenue lors de la première analyse et donc appartenant au Run 1, et une deuxième mesure appartenant au Run 2 puisqu'ayant été effectué en même temps que l'analyse des échantillons du jour 35. On obtient donc un modèle à mesures répétées à trois facteurs fixes (le groupe, le jour et le run) avec des répétitions sur les sujets suivant le facteur run. En considérant l'effet run comme étant fixe, on obtiendra une valeur numérique lors de l'analyse de ce modèle représentant une valeur moyenne de différence entre les deux runs ce qui nous servira afin de corriger les titres. Cependant, les variables groupe et jour

sont totalement indépendants de la variable run ce qui implique que les résultats d'analyse obtenus sur l'effet run, c'est-à-dire ce qui nous intéresse, seront identiques que l'on enlève un de ces effets, ou les deux, et/ou leur interaction ou qu'on les laisse. J'ai donc procédé à une sélection manuelle de modèle en utilisant comme critère de sélection les p-valeurs et en partant du modèle suivant:

(Voir Annexe 3 Fig.3,4,5 et 6)

$$Y_{ijk_s} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_k + \gamma_s + \varepsilon_{ijk_s}, \quad \text{avec } 1 \leq i \leq 2, 1 \leq j \leq 2, 1 \leq k \leq 2 \text{ et } 1 \leq s \leq 29.$$

Cette sélection n'aura donc pas modifié les résultats de l'effet run. Pour étudier ce problème posé sur l'effet run, on a donc choisi le modèle suivant:

$$(2) \quad Y_{iks} = \mu + \alpha_i + \delta_k + \gamma_s + \varepsilon_{iks}, \quad \text{avec } 1 \leq i \leq 2, 1 \leq k \leq 2, 1 \leq s \leq 29.$$

où α_i représente l'effet fixe du i-ème temps d'étude.

δ_k représente l'effet fixe du run k.

γ_s représente l'effet aléatoire du s-ème sujet.

ε_{iks} l'erreur résiduelle.

On impose ici comme contrainte que :

$$\alpha_2 = 0 ; \delta_2 = 0 ; \gamma_i \sim N(0, \sigma_\gamma^2) ; \varepsilon_{iks} \sim N(0, \sigma^2)$$

J'ai également vérifié les hypothèses sur les résidus. Les résultats sur ce modèle-ci sont disponibles en Annexe 4.

Enfin un troisième modèle a été utilisé afin d'extraire la variabilité de l'effet run. En effet, en partant de l'ensemble des données il a fallu estimer cette variabilité mais sans oubliant le fait que cet effet est lié à l'effet jour. On s'est donc basé sur ce modèle :

$$(3) \quad Y_{ijk_s} = \mu + \alpha_i + \beta_j + \delta_k + (\alpha\delta)_{ik} + \gamma_s + \varepsilon_{ijk_s}, \quad \text{avec } 1 \leq i \leq 3, 1 \leq j \leq 2, 1 \leq k \leq 2 \text{ et } 1 \leq s \leq 130.$$

où Y_{ijk_s} représente le $\log_{10}(\text{HI})$ avec les HI non-corrigés

α_i représente l'effet fixe du i-ème temps d'étude

δ_k représente l'effet aléatoire du run k.

β_j représente l'effet fixe du j-ème groupe (~ vaccin)

γ_s représente l'effet aléatoire du s-ème sujet

ε_{ijk_s} l'erreur résiduelle.

On impose ici comme contrainte que :

$$\alpha_3 = 0 ; \beta_2 = 0 ; \delta_k \sim N(0, \sigma_\delta^2) ; (\alpha\delta)_{ik} \sim N(0, \sigma_{\alpha\delta}^2) ; \gamma_i \sim N(0, \sigma_\gamma^2) ; \varepsilon_{iks} \sim N(0, \sigma^2)$$

L'interaction entre le jour et le run permet de réduire l'erreur sur la variabilité de l'effet run. En effet, comme notre plan d'étude n'est pas complet, le jour et le run sont fortement corrélés. On peut ainsi supposer que le résultat obtenu est approximativement celui que l'on aurait eu si le plan avait été complet et qu'ainsi la corrélation entre l'effet jour et l'effet run aurait été nulle comme pour les données du bridge. Afin d'obtenir cette variabilité inter-run que nous avons besoin pour le calcul de covariance (voir Annexe 7), il a donc fallu utiliser ce modèle.

3.2.6. Informations préliminaires

- L'orientation initiale s'est portée sur la modélisation des cinétiques individuelles afin de travailler avec des titres estimés. Il a donc fallu que je trouve une fonction de trois paramètres à estimer au maximum, puisque n'ayant au minimum que trois titres par sujet. Je me suis aidé d'une publication de F.Tibaldi et al.(2009)¹ où l'on peut trouver une fonction non-linéaire à quatre paramètres servant à modéliser la cinétique de données immunologique.

La fonction décrite dans ce document qui m'a servi de référence est la suivante:

$$f(t,\beta) = \frac{\beta_1(1+\beta_2).e^{-\beta_3.t}}{1+\beta_2.e^{-\beta_4.t}}$$

où $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ est le vecteur des paramètres estimer.

β_1 correspond à la valeur du titre au temps $t=0$.

β_2 est lié à la valeur maximale lorsque les autres paramètres sont fixes.

β_3 représente le taux d'élimination, donc de décroissance de la courbe.

β_4 détermine la pente d'accroissement.

J'ai donc choisi de fixer β_3 à 0, en admettant qu'au Jour21 on n'avait pas encore atteint le maximum des titres avec une dose, et donc, qu'on n'était pas encore en phase décroissante. De même, j'ai fait la même supposition au Jour 35. Cependant, lors de la modélisation à l'aide de la procédure SAS PROC NLIN, l'ajustement individuel était difficile pour certains sujets. Il fonctionne cependant relativement bien sur l'évolution des moyennes géométriques de ces titres.

Cette partie de mon travail, partie sur laquelle je n'irai pas plus loin puisque n'ayant pas réellement aboutit, m'a servi à mieux visualiser l'évolution des titres, les problèmes existants et m'a ainsi apporter une vision globale de l'étude sur laquelle j'allais travailler (Voir Annexe 2).

En conséquence, le travail s'est alors focalisé sur l'idée d'une correction. Deux approches ont été envisagées, une première entièrement fréquentiste et une deuxième utilisant de l'analyse bayésienne sur l'estimation de la moyenne et de la variance de l'effet run. Cependant, seule l'approche fréquentiste sera détaillée. En effet, par manque de temps, la méthode bayésienne sera à finaliser ultérieurement.

- Lors de ces recherches, j'ai été amené à devoir estimer le logarithme de base 10 des moyennes géométriques de ces titres HI, moyennes géométriques que l'on va appeler GMT, pour chacun des trois jours et deux groupes. Cela revient en fait à estimer au mieux les moyennes arithmétiques des $\log_{10}(\text{HI})$. En effet,

$$\log_{10}(\text{GMT}) = \log_{10}\left(\sqrt[n]{\prod_{i=1}^n \text{HI}_i}\right) = (1/n) \cdot \sum_{i=1}^n \log_{10}(\text{HI})_i = \overline{\log_{10}(\text{HI})}$$

1. F.Tibaldi, E.Ledent and R.Van der Most. A Statistical Approach Using Non-Linear Modeling for the Analysis of Immunological Data, 2009.

- Voici les valeurs des $\log_{10}(\text{GMT})$ non corrigés (Voir Annexe 5 Fig.1) :

[log₁₀\(GMT\) estimés avant correction.](#)

	Jour 0	Jour 21	Jour 35
Groupe 1	0.9364	2.5551	2.6374
Groupe 2	1	2.6170	2.5052

On remarque bien que pour le Groupe 1 l'augmentation entre le Jour 21 et le Jour 35 est légère et pour le Groupe 2 il y'a une diminution des titres, ce qui pose naturellement problème.

IV. Développement et solutions apportées

4.1. Approche fréquentiste

4.1.1. Idée générale de l'approche

L'approche fréquentiste est une approche basée sur la correction étude par étude. On ne tient aucunement compte du passé ni des résultats d'autres études similaires terminées ou en cours. L'idée est simple, se servir des données du bridge afin d'estimer un écart moyen entre le Run 1 et le Run 2, appelé écart inter-run, puis de corriger les titres afin que la variabilité liée à cet effet soit éliminée.

Soit $Y_{t,i,k,j}$ la réponse au temps t du sujet i du groupe k pour le run j

Alors, posons $\delta_i = Y_{t,i,k,1} - Y_{t,i,k,2}$ pour $1 \leq i \leq 29$, l'écart inter-run du i -ème individu du bridge. Donc on obtient (Voir Annexe5 Fig2),

$$\begin{aligned}\bar{\delta} &= (1/29) \cdot \sum_{i=1}^{29} \delta_i \\ &= 0.13543\end{aligned}$$

Cependant une question s'est naturellement posée à nous, de quelles manières allons-nous corriger ces titres? Plusieurs possibilités ont ainsi été avancées.

4.1.2. Corrections des titres

Trois possibilités de retrait de cet effet ont été étudiées. La première consiste à soustraire $\bar{\delta}/2$ à tous les titres du Run1 et à rajouter $\bar{\delta}/2$ à tous les titres du Run2 y compris aux titres du bridge. La deuxième a été de soustraire à tous les titres du Run1 $\bar{\delta}$ et la troisième était d'ajouter $\bar{\delta}$ à tous les titres du Run2. A noter que lorsque sur une journée et un sujet il y avait la présence d'un bridging on sélectionnait le milieu des deux observations comme étant l'observation à sélectionner. Cependant, en prenant ceci au cas par cas on se rend compte que l'on corrigera soit de trop soit pas assez chaque sujet. Mais cela ne pose aucunement problème. En effet, l'intérêt de l'étude n'est pas porté sur l'étude sujet par sujet mais sur une approche globale, il n'était donc pas nécessaire de corriger chaque sujet présent dans le bridge avec son propre écart inter-run tant qu'en moyenne cet écart était effacé.

Ces trois corrections ont été faites et les résultats comparés. Pour cela, on considère des contrastes à l'intérieur de la procédure SAS PROC MIXED de la forme 1 -1. Posons $\bar{Y}'_{t,..,k}$ la moyenne arithmétique des titres corrigés. On a alors que:

$$\bar{Y}'_{t,..,k} - \bar{Y}'_{t',,..,k} = \log_{10}(\text{GMT}_{t,k}) - \log_{10}(\text{GMT}_{t',k}) = \log_{10}\left(\frac{\text{GMT}_{t,k}}{\text{GMT}_{t',k}}\right),$$

avec $\text{GMT}_{t,k}$ le GMT au jour t pour le groupe k .

$$\Rightarrow \text{GMR}_{k, \frac{t}{t'}} = \frac{\text{GMT}_{t,k}}{\text{GMT}_{t',k}} = 10^{\log_{10}\left(\frac{\text{GMT}_{t,k}}{\text{GMT}_{t',k}}\right)}$$

Il s'avère qu'au niveau des rapports de GMT, ils restent similaires lorsqu'on change de rectification. En effet, quelque soit la correction apportée, cette correction va naturellement s'effacer en faisant la différence des moyennes à deux temps différents. Une démonstration pour un cas particulier est donnée en annexe. On en conclut donc que ces trois corrections sont équivalentes au niveau du rapport des GMT (Voir Annexe 6). En effet, suivant la correction choisie, les valeurs des GMT vont être modifiées mais leur rapport reste exactement le même ce qui est la base de notre étude. L'objectif n'est pas de connaître parfaitement la valeur des GMT mais bien de pouvoir les comparer correctement en effectuant leur rapport les uns par rapport aux autres. Après concertation avec le personnel concerné, on a donc décidé de choisir la première méthode comme correction puisqu'il n'y a pas de raison de ne corriger qu'un seul Run ou de corriger un Run plus que l'autre. En effet, ne connaissant pas d'a priori sur les causes ou d'indications nous permettant d'affirmer que les valeurs obtenues au Run 2 sont moins bonnes que celles obtenues au Run1, on a décidé de corriger chacun des deux Run de façon égale.

On obtient ainsi avec cette correction choisie les six $\log_{10}(\text{GMT})$ corrigés suivant (Voir Annexe5 Fig.3):

$\log_{10}(\text{GMT})$ estimés après Correction.

	Jour 0	Jour 21	Jour 35
Groupe 1	0.8769	2.4871	2.7051
Groupe 2	0.9345	2.5461	2.5729

Ces résultats, devenus désormais cohérents, sont quasiment similaires aux différents rapports obtenus au Run 3.

$\log_{10}(\text{GMT})$ estimés au Run3.

	Jour 0	Jour 21	Jour 35
Groupe 1	1.0156	2.6937	2.8636
Groupe 2	1.0908	2.7422	2.7446

Rapports des $\log_{10}(\text{GMT})$ estimés.

		$\frac{\text{Jour 21}}{\text{Jour 0}}$	$\frac{\text{Jour 35}}{\text{Jour 0}}$	$\frac{\text{Jour 35}}{\text{Jour 21}}$
Groupe 1	Run 1&2	2.8362	3.0848	1.0877
Groupe 1	Run 3	2.6523	2.8196	1.0631
Groupe 2	Run 1&2	2.7246	2.7532	1.0105
Groupe 2	Run 3	2.5139	2.5161	1.0009

4.1.3. Estimation des variances et intervalles de confiance

Il s'agissait ensuite d'estimer la variance des nouveaux $\log_{10}(\text{GMT})$ corrigés en tenant compte, notamment, de la variance de la correction. Une fois cette variance estimée, il ne restait plus qu'à créer le nouvel intervalle de confiance autour de chacun des $\log_{10}(\text{GMT})$:

Soient $Y_i \stackrel{iid}{\sim} N(m, \sigma^2)$, alors on sait que l'estimateur empirique de la moyenne m , que l'on va noter \bar{Y} , suit une $N(m, \frac{\sigma^2}{n})$.

D'où $\frac{\bar{Y}-m}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$.

Soit S^2 l'estimateur non biaisé de σ^2 , $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$. Donc on a que :

$$(n-1) \cdot \frac{S^2}{\sigma^2} = \sum_{i=1}^n \frac{(Y_i - \bar{Y})^2}{\sigma^2}$$

Or d'après le théorème de Cochran, on sait que : $(n-1) \cdot \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$.

Donc, si on pose $X = \frac{\bar{Y}-m}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{Y}-m}{\frac{S}{\sqrt{n}}} / \sqrt{\frac{S^2}{\sigma^2}}$.

On en conclut donc que $X \sim \frac{N(0,1)}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}}$ ie $X \sim T_{n-1}$.

On peut donc déterminer un intervalle de confiance à 95% pour m :

Soit $q_{n-1}^{p\%}$ le quantile d'une loi de Student à $n-1$ degrés de liberté au seuil de $p\%$.

Alors on a par définition que $P(|X| \leq q_{n-1}^{95\%}) = 0.95$

$$\text{Or } |X| \leq q_{n-1}^{95\%} \Leftrightarrow \left| \frac{\bar{Y}-m}{\frac{S}{\sqrt{n}}} \right| \leq q_{n-1}^{95\%} \Leftrightarrow q_{n-1}^{97.5\%} \leq \frac{\bar{Y}-m}{\frac{S}{\sqrt{n}}} \leq q_{n-1}^{97.5\%} \Leftrightarrow m \in \left[\bar{Y} - q_{n-1}^{97.5\%} \cdot \frac{S}{\sqrt{n}} ; \bar{Y} + q_{n-1}^{97.5\%} \cdot \frac{S}{\sqrt{n}} \right]$$

Cependant, lorsqu'on est en présence d'un effectif suffisamment grand, on peut se permettre d'approximer notre quantile de loi de Student par un quantile de loi Normal $N(0,1)$. Cependant, étant donné le faible effectif de sujet par jour et par groupe, on a préféré ne pas faire cette approximation et ainsi calculer le quantile de la loi de Student correspondant pour construire l'intervalle de confiance recherché.

Il s'est alors posé le problème des degrés de liberté. En effet, dans notre cas, les $\log_{10}(\text{GMT})$ corrigés ne peuvent pas être considérés comme étant des moyennes fixes. En effet, on a une variabilité de ces moyennes dû à la variabilité de la correction δ . Il n'est donc pas possible de calculer la variance de ces $\log_{10}(\text{GMT})$ corrigés de façon empirique comme décrit ci-dessus. Pour calculer ces variances, il faut la décomposer :

$$\begin{aligned} \text{Var}(\overline{Y'_{t,..,k}}) &= \text{Var}(\overline{Y_{t,..,k,j}}) + \text{Var}(\overline{\delta/2}) \pm 2 \cdot \text{Cov}(\overline{Y_{t,..,k,j}}, \overline{\delta/2}) \\ &= \text{Var}(\overline{Y_{t,..,k,j}}) + \text{Var}(\overline{\delta})/4 \pm \text{Cov}(\overline{Y_{t,..,k,j}}, \overline{\delta}) \end{aligned}$$

Nous sommes donc dans le cas d'une combinaison linéaire de deux variances associées à deux variables corrélées. Cette fois, les deux variances obtenues peuvent être estimées de façon empirique sans aucun problème. Le seul problème résulte dans la notion de degré de liberté. En effet, dans l'explication faite ci-dessus le degré de liberté associé au quantile n'est rien d'autre que $n-1$, ceci étant dû au fait que l'on estimait la variance de la variable cible de façon empirique grâce à sa

forme non biaisée. Ici, on ne peut pas procéder de la même façon étant donné que notre variance est composée de deux variances que l'on va estimer chacune de façon empirique mais chacune avec un effectif différent.

Il fallait donc trouver un moyen d'estimer le degré de liberté associé à ces $\log_{10}(\text{GMT})$ corrigés. Pour cela, j'ai utilisé la formule de Welch-Satterthwaite adaptée aux cas corrélés qui permet d'obtenir une approximation des degrés de liberté d'une combinaison linéaire de variances. J'ai trouvé cette formule ainsi que les explications et la démonstration dans un article du Dr. Howard Castrup.²

Soit $\nu_{t,k}$ le degré de liberté recherché pour le jour t du groupe k , alors d'après l'extension de l'équation de Welch-Satterthwaite aux cas non-corrélés on a :

$$\nu_{t,k} \simeq \frac{(\sigma_T^2)^2}{\sum_{i=1}^n \frac{(\sigma_i^2)^2}{\nu_i} + \sum_{i=1}^{n-1} \sum_{j>i}^n \rho_{ij}^2 \cdot \left(\frac{\sigma_i^2}{\nu_i}\right) \cdot \left(\frac{\sigma_j^2}{\nu_j}\right) \cdot (\nu_i + \nu_j + 1/2) + 2 \cdot \sum_{i=1}^{n-1} \sum_{j>i}^n \rho_{ij} \cdot \sigma_i \cdot \sigma_j \cdot \left(\frac{\sigma_i^2}{\nu_i} + \frac{\sigma_j^2}{\nu_j}\right)}$$

avec :

- σ_T^2 la variance totale de $\overline{Y'_{t,..,k}}$.
- σ_i^2 la i -ème variance de la combinaison linéaire.
- ν_i le degré de liberté associé à la i -ème variance.
- ρ_{ij} le coefficient de corrélation entre la i -ème et j -ème variance.
- $n = 2$ dans notre étude.

Comme dit précédemment, il est facile de calculer la variance des $\overline{Y_{t,..,k,j}}$, ainsi que la variance de $\bar{\delta}$ en les estimant de façon empirique. L'estimation de la covariance $\text{Cov}(\overline{Y_{t,..,k,j}}, \bar{\delta})$ est plus compliquée à obtenir. Une hypothèse a été faite à ce niveau là. En effet, nous nous sommes servis d'un modèle complet entièrement simulé pour vérifier la valeur des différentes covariances qui composent $\text{Cov}(\overline{Y_{t,..,k,j}}, \bar{\delta})$. On a ainsi trouvé des relations assez simples entre ces différentes covariances et les variabilité inter-run, inter-sujet et résiduelles. L'hypothèse que nous avons faite est donc de supposer que nos résultats, vrais pour un modèle complet, restent vrais pour notre modèle. Par manque de temps, il ne m'a pas été possible de chercher à démontrer ces résultats obtenus ni de chercher à démontrer notre hypothèse (Voir Annexe 7).

2. Dr. Howard Castrup. A Welch-Satterthwaite Relation for Correlated Errors [en ligne]. Measurement Science Conference, Pasadena, CA, Mars 2010. Disponible sur : < http://www.isgmax.com/Articles_Papers/Welch-Satterthwaite%20for%20Correlated%20Errors.pdf >

4.1.4. Résultats

J'ai donc fini par trouver les résultats suivants :
(Voir la signification des nouvelles notations dans l'Annexe 7)

Groupe 1 :

$$\begin{aligned}\text{Cov}(\overline{Y_{0,..,1,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{n_{1,1}}{(n_1+n_2) \cdot N_1} \cdot \sigma_{Res}^2 \\ \text{Cov}(\overline{Y_{21,..,1,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{n_{1,2}}{(n_1+n_2) \cdot N_1} \cdot \sigma_{Res}^2 \\ \text{Cov}(\overline{Y_{35,..,1,2}}, \overline{\delta.}) &= -\sigma_{B-Run}^2\end{aligned}$$

Groupe 2 :

$$\begin{aligned}\text{Cov}(\overline{Y_{0,..,2,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{n_{2,1}}{(n_2+n_1) \cdot N_2} \cdot \sigma_{Res}^2 \\ \text{Cov}(\overline{Y_{21,..,2,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{n_{2,2}}{(n_2+n_1) \cdot N_2} \cdot \sigma_{Res}^2 \\ \text{Cov}(\overline{Y_{35,..,2,2}}, \overline{\delta.}) &= -\sigma_{B-Run}^2\end{aligned}$$

avec σ_{B-Run}^2 la variabilité inter-run et σ_{Res}^2 la variabilité résiduelle.

Donc pour le groupe 1 :

On a 64 sujets dans ce groupe, dont 22 appartiennent également au bridge effectué sur les 40 sujets de départ, dont seuls 18 ont été retenus, 12 ont été bridgé au jour 21 et 6 au jour 0.

D'où $N_1 = 64$, $n_1 = 18$, $n_{1,2} = 12$ et, par souci de clarté, on suppose avoir classé l'échantillon des éléments du bridge de tel sorte que les éléments du Groupe 1 soient placés en premier avant les éléments du Groupe 2 et au sein de chaque groupe les éléments du Jour 0 avant les éléments du Jour 21. On a $n_1 + n_2 = 29$ et on déduit que $n_{1,1} = 6$. Donc on obtient :

$$\begin{aligned}\text{Cov}(\overline{Y_{0,..,1,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{6}{29 \cdot 64} \cdot \sigma_{Res}^2 = \sigma_{B-Run}^2 + \frac{3}{928} \cdot \sigma_{Res}^2 \\ \text{Cov}(\overline{Y_{21,..,1,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{12}{29 \cdot 64} \cdot \sigma_{Res}^2 = \sigma_{B-Run}^2 + \frac{6}{928} \cdot \sigma_{Res}^2 \\ \text{Cov}(\overline{Y_{35,..,1,2}}, \overline{\delta.}) &= -\sigma_{B-Run}^2\end{aligned}$$

Et pour le groupe 2 :

On a 66 sujets dans ce groupe, dont 18 appartiennent également au bridge effectué sur les 40 sujets de départ, dont seuls 11 ont été retenus, 8 ont été bridgé au jour 21 et 3 au jour 0.

D'où $N_2 = 66$, $n_2 = 11$, $n_{2,1} = 3$, $n_{2,2} = 8$. On suppose à nouveau avoir ordonnés les données comme décrit ci-dessus pour le Groupe 1, mais pour le Groupe 2. Donc on obtient :

$$\begin{aligned}\text{Cov}(\overline{Y_{0,..,2,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{3}{29 \cdot 66} \cdot \sigma_{Res}^2 \\ \text{Cov}(\overline{Y_{21,..,2,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{8}{29 \cdot 66} \cdot \sigma_{Res}^2 = \sigma_{B-Run}^2 + \frac{4}{957} \cdot \sigma_{Res}^2 \\ \text{Cov}(\overline{Y_{35,..,2,2}}, \overline{\delta.}) &= -\sigma_{B-Run}^2\end{aligned}$$

Voici donc les résultats numérique pour chacun des groupes et jours (Voir Annexe 8 et 9):

La variabilité de l'effet run a été estimée à l'aide du modèle (3) et la variabilité résiduelle a elle été estimée à partir du modèle (2) en supposant que celle-ci serait identique si l'on avait eu un plan d'étude complet. C'est-à-dire, si l'on avait un plan similaire au plan d'étude du bridge mais sur tous les sujets ainsi que sur les trois temps d'études.

• **Groupe 1 Jour 0 :**

$$\begin{aligned} \text{Var}(\overline{Y_{0,..,1,1}}) &\simeq 0.002853 \\ \text{Var}(\overline{\delta.})/4 &\simeq 0.000522/4 = 0.0001305 \\ \text{Cov}(\overline{Y_{0,..,1,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{3}{928} \cdot \sigma_{Res}^2 = 0.000976 + \frac{3}{928} \times 0.007570 \simeq 0.001 \\ \text{Var}(Y'_{0,..,1}) &\simeq 0.002853 + 0.0001305 - 0.001 = 0.0019828362 \end{aligned}$$

• **Groupe 1 Jour 21 :**

$$\begin{aligned} \text{Var}(\overline{Y_{21,..,1,1}}) &\simeq 0.003873 \\ \text{Var}(\overline{\delta.})/4 &\simeq 0.000522/4 = 0.0001305 \\ \text{Cov}(\overline{Y_{21,..,1,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{6}{928} \cdot \sigma_{Res}^2 = 0.000976 + \frac{6}{928} \times 0.007570 \simeq 0.001025 \\ \text{Var}(Y'_{21,..,1}) &\simeq 0.003873 + 0.0001305 - 0.001025 = 0.0029781745 \end{aligned}$$

• **Groupe 1 Jour 35 :**

$$\begin{aligned} \text{Var}(\overline{Y_{35,..,1,2}}) &\simeq 0.002721 \\ \text{Var}(\overline{\delta.})/4 &\simeq 0.000522/4 = 0.0001305 \\ \text{Cov}(\overline{Y_{35,..,1,2}}, \overline{\delta.}) &= -\sigma_{B-Run}^2 = -0.000976 \\ \text{Var}(Y'_{35,..,1}) &\simeq 0.002721 + 0.0001305 - 0.000976 = 0.0018756623 \end{aligned}$$

• **Groupe 2 Jour 0 :**

$$\begin{aligned} \text{Var}(\overline{Y_{0,..,2,1}}) &\simeq 0.002766 \\ \text{Var}(\overline{\delta.})/4 &\simeq 0.000522/4 = 0.0001305 \\ \text{Cov}(\overline{Y_{0,..,2,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{3}{29.66} \cdot \sigma_{Res}^2 = 0.000976 + \frac{1}{638} \times 0.007570 \simeq 0.000988 \\ \text{Var}(Y'_{0,..,2}) &\simeq 0.002766 + 0.0001305 - 0.000988 = 0.0019089948 \end{aligned}$$

• **Groupe 2 Jour 21 :**

$$\begin{aligned} \text{Var}(\overline{Y_{21,..,2,1}}) &\simeq 0.003651 \\ \text{Var}(\overline{\delta.})/4 &\simeq 0.000522/4 = 0.0001305 \\ \text{Cov}(\overline{Y_{21,..,2,1}}, \overline{\delta.}) &= \sigma_{B-Run}^2 + \frac{8}{29.66} \cdot \sigma_{Res}^2 = 0.000976 + \frac{4}{957} \times 0.007570 \simeq 0.001008 \\ \text{Var}(Y'_{21,..,2}) &\simeq 0.003651 + 0.0001305 - 0.001008 = 0.0027736173 \end{aligned}$$

• **Groupe 2 Jour 35 :**

$$\begin{aligned} \text{Var}(\overline{Y_{35,..,2,2}}) &\simeq 0.002543 \\ \text{Var}(\overline{\delta.})/4 &\simeq 0.000522/4 = 0.0001305 \\ \text{Cov}(\overline{Y_{35,..,2,2}}, \overline{\delta.}) &= -\sigma_{B-Run}^2 = -0.000976 \\ \text{Var}(Y'_{35,..,2}) &\simeq 0.002543 + 0.0001305 - 0.000976 = 0.0016977803 \end{aligned}$$

Tableau récapitulatif des nouvelles variances:

	Jour 0	Jour 21	Jour 35
Groupe 1	0.0019828362	0.0029781745	0.0018756623
Groupe 2	0.0019089948	0.0027736173	0.0016977803

Ainsi il devient également possible de calculer $\rho_{t,k}$ en posant :

$$\rho_{t,k} = \frac{Cov(\overline{Y_{t,..,k,j}}, \overline{\delta/2})}{\sqrt{Var(\overline{Y_{t,..,k,j}})} \cdot \sqrt{Var(\overline{\delta/2})}} = \frac{Cov(\overline{Y_{t,..,k,j}}, \overline{\delta})}{\sqrt{Var(\overline{Y_{t,..,k,j}})} \cdot \sqrt{Var(\overline{\delta})}}$$

D'où : $\rho_{0,1} = 0.8197892352$; $\rho_{21,1} = 0.7208268225$; $\rho_{35,1} = -0.81885354$
 $\rho_{0,2} = 0.8220097126$; $\rho_{21,2} = 0.7298730971$; $\rho_{35,2} = -0.847005936$

(Voir Annexel1 pour tous les résultats ci-dessus ainsi que ceux à venir)

On a donc tous les éléments afin de calculer les degrés de libertés nécessaires.

A l'aide de SAS, on obtient donc:

$$\nu_{0,1} = 38.51599 ; \nu_{21,1} = 53.74568 ; \nu_{35,1} = 80.76914$$

$$\nu_{0,2} = 37.56873 ; \nu_{21,2} = 51.54307 ; \nu_{35,2} = 76.66523$$

Il ne reste donc plus qu'à déterminer les intervalles de confiance autour des différents $\log_{10}(\text{GMT})$:

Pour cela, voici tout d'abord les différents quantiles de la loi de Student obtenus :

$$q_{\nu_{0,1}}^{0.975} = 2.0235 ; q_{\nu_{21,1}}^{0.975} = 2.0051 ; q_{\nu_{35,1}}^{0.975} = 1.98977$$

$$q_{\nu_{0,2}}^{0.975} = 2.02516 ; q_{\nu_{21,2}}^{0.975} = 2.00707 ; q_{\nu_{35,2}}^{0.975} = 1.99139$$

Soit $m_{t,k}$ le $\log_{10}(\text{GMT})$ réel au temps t du groupe k. Alors on obtient :

$$m_{0,1} \in [\overline{Y'_{0,..,1}} - q_{\nu_{0,1}}^{0.975} \cdot \frac{S}{\sqrt{n}} ; \overline{Y'_{0,..,1}} + q_{\nu_{0,1}}^{0.975} \cdot \frac{S}{\sqrt{n}}]$$

$$\Leftrightarrow m_{0,1} \in [\overline{Y'_{0,..,1}} - q_{\nu_{0,1}}^{0.975} \cdot \sqrt{Var(Y'_{0,..,1})} ; \overline{Y'_{0,..,1}} + q_{\nu_{0,1}}^{0.975} \cdot \sqrt{Var(Y'_{0,..,1})}]$$

$$\Leftrightarrow m_{0,1} \in [0.8769 - 2.0235039151 \times 0.0534115 ; 0.8769 + 2.0235039151 \times 0.0534115]$$

$$\Leftrightarrow m_{0,1} \in [0.7867734518 ; 0.9669828615]$$

Puis pour les autres, de la même manière on obtient :

$$m_{21,1} \in [2.3776825551 ; 2.5965294368] ; m_{35,1} \in [2.6188938755 ; 2.7912436646]$$

$$m_{0,2} \in [0.8460301013 ; 1.0229966575] ; m_{21,2} \in [2.4403534473 ; 2.6517587403]$$

$$m_{35,2} \in [2.490826017 ; 2.6549332602]$$

Finalement, il ne reste plus qu'à tous reconvertir en terme de HI et non plus en $\log_{10}(\text{HI})$. Pour cela, il suffit de modifier tous les résultats x en 10^x .

On obtient donc finalement :

GMT estimés après Correction.

	Jour 0	Jour 21	Jour 35
Groupe 1	7.5314	306.98	507.07
Groupe 2	8.6	351.61	374.01

$$m_{0,1} \in [6.1203104462 ; 9.267932487] ; m_{21,1} \in [238.60665644 ; 394.93846767]$$

$$m_{35,1} \in [415.80899074 ; 618.36324077]$$

$$m_{0,2} \in [7.0150391848 ; 10.543787814] ; m_{21,2} \in [275.64711238 ; 448.49617168]$$

$$m_{35,2} \in [309.61786888 ; 451.78651122]$$

(Voir Graphiques Annexe 12)

N'oublions pas que nous nous intéressions à la cinétique. J'ai donc cherché à calculer la variance, pour les deux groupes, de la différence des moyennes entre le jour 35 et 21. Les résultats sont détaillés en Annexe 9 et montrent les limites de notre méthode.

4.2. Approche bayésienne

Une autre approche afin de résoudre ce problème aurait été d'utiliser une méthode bayésienne afin d'estimer les paramètres de la loi de l'effet run.

4.2.1. Motivation

Nous avons comme apriori qu'en moyenne l'écart inter-run était nul. L'idée était donc de partir, comme loi à priori, d'une loi normale de moyenne 0 et de variance très grande afin que celle-ci soit non-informative. Il fallait ensuite déterminer la loi à postérieure de l'effet run puis recommencer ce processus avec une autre étude, réalisée de façon similaire, avec comme loi à priori la loi à postérieure obtenue précédemment. Il aurait ensuite fallu corriger les études à l'aide des paramètres des lois à postérieures obtenues. L'idée était ainsi d'arriver, asymptotiquement, à une forte précision de la variance de l'effet run qui devrait posséder une moyenne de 0. Il suffirait donc, à terme, de ne corriger que la variance des moyennes des GMT et plus les titres eux-mêmes. Il s'agit là d'une approche qui paraît, mathématiquement, plus précise que l'approche fréquentiste réalisée ci-dessus.

4.2.2. Inconvénient

L'idée d'arriver à terme à une non-corrrection des titres mais uniquement de la variance des moyennes et donc d'une correction uniquement de l'intervalle de confiance pose un problème plus visuel et d'interprétation possible dans le monde professionnel. Ceci n'est pas négligeable lorsqu'il s'agit de présenter ou soumettre des résultats à un certain auditoire ou conseil.

Cependant, par manque de temps, il ne m'a pas été possible de mener à terme cette approche. J'ai seulement pu entamer un début de code SAS pour pouvoir observer l'évolution sur trois études à l'aide de la procédure SAS, PROC GENMOD.

V. Conclusion

5.1. Questionnements au sujet des résultats

Les résultats obtenus lors de notre analyse sont soumis à certaines vérifications et démonstrations qu'il ne m'a pas été possible d'effectuer compte tenu de la durée du stage. Ainsi, le fait que les variances des GMT avec rectification de variance soient plus faibles que celles des GMT sans rectification de variance est un résultat surprenant mais naturellement possible. Malgré le fait qu'il s'agirait là d'un résultat plutôt bénéfique et très satisfaisant, il faudrait tout d'abord démontrer les quelques suppositions effectuées au cours de notre étude. La validité des modèles statistiques utilisés ainsi que les bases de données utilisées afin de calculer les différentes variabilités seraient également à révéifier. En effet, à la vue des données que nous possédions sur cette étude, il serait intéressant de vérifier que les résultats théoriques que nous avons utilisés, vrais sur un plan complet, le restent sur notre étude en question. On remarque d'ailleurs, dans l'Annexe 9 que l'on obtient pour le groupe 2 une variance négative ce qui est révélateur des problèmes qui existent encore au sein du développement. Une solution qui réglerait ces problèmes de variances négatives auraient été de supposer l'indépendance entre les GMT non corrigés et l'écart inter-run afin d'avoir une covariance nulle et ainsi être sûr d'obtenir des variances positives. Ceci réduit de façon significative les problèmes rencontrés.

Il serait également intéressant de poursuivre l'approche bayésienne qui serait certainement moins approximative puisque ne nécessitant pas, à priori, d'hypothèses comme celles que nous avons faites à travers l'approche fréquentiste.

5.2. Bilan général du stage

Ce stage a été une très bonne expérience pour moi sur plusieurs plans. Tout d'abord, d'un point de vue professionnel naturellement, ce stage m'a permis de réellement réaliser ce qu'était le travail en entreprise. Celui-ci étant totalement différent du travail fourni à l'université, j'ai appris à mieux gérer mon temps, à travailler de façon efficace, concise et synthétique mais précise. La collaboration entre personnes et entre services au sein d'une entreprise est un aspect fondamental que je ne soupçonnais pas forcément mais qui est extrêmement utile afin de pouvoir prendre du recul sur les recherches que l'on fait et avoir un point de vue extérieur, neuf. Cela m'a souvent permis de résoudre les problèmes que j'ai rencontrés.

En effet, ces recherches ont souvent été compliquées par la nature peu évidente des données que j'ai eu à traiter. J'ai souvent résolu ces problèmes à l'aide d'avis extérieurs qui m'apportaient une autre vision du problème et qui me permettaient ainsi de voir le problème, et donc la solution, sous un autre angle.

Sur un point de vue mathématique et statistique, ce stage m'a également beaucoup apporté et beaucoup appris. En effet, de la même manière que la façon de travailler est différente entre l'université et le monde de l'entreprise, la manière de faire des statistiques l'est également. La principale différence que j'ai remarqué est qu'une fois entré dans les statistiques basées sur des études réelles, tout ne se passe pas toujours aussi bien que l'on pouvait l'espérer. Par exemple, lors de la vérification des conditions d'applications des méthodes utilisées ou encore sur la gestion des données manquantes (censures, troncatures...). Malgré ces problèmes, il faut alors trouver d'autres solutions, alternatives aux méthodes classiques afin de traiter au mieux les données que

l'on possède. Il est assez rare que l'on puisse laisser tomber totalement une étude réalisée sous le prétexte que certaines conditions ne sont pas remplies ou que quelques données posent problèmes compte tenu de la difficulté de refaire une étude et de l'aspect financier et commercial mis en jeu. Tout ceci, naturellement, en restant dans le cadre de statistiques parfaitement correctes.

Enfin, il s'agissait ici d'une étude très complexe et ambitieuse au vue du temps que je possédais. En effet, notre base de données étant relativement complexe, il aurait fallu tenter d'autres méthodes et vérifier plus d'hypothèses que ce que je n'ai eu le temps de faire. Il s'agissait ici d'un réel sujet de recherche auquel j'ai pu apporter ma participation, bien que n'ayant pu y trouver une solution sûr et parfaitement stable. Cela reste une très bonne expérience et un stage qui m'a beaucoup appris durant ces trois mois.

Annexes

N.B. : Les annexes contenant les graphiques ainsi que les tableaux de sortie SAS sont manquantes sur ce PDF.

• **Annexe 6 : Données de Comparaison des trois types de corrections des $\log_{10}(GMT)$.**

	$\overline{Y'_{35,..,1}} - \overline{Y'_{21,..,1}}$	$\overline{Y'_{35,..,2}} - \overline{Y'_{21,..,2}}$	$\overline{Y'_{21,..,1}} - \overline{Y'_{0,..,1}}$	$\overline{Y'_{21,..,2}} - \overline{Y'_{0,..,2}}$	$\overline{Y'_{35,..,1}} - \overline{Y'_{0,..,1}}$	$\overline{Y'_{35,..,2}} - \overline{Y'_{0,..,2}}$
Run1- δ	1.651820205	1.063710742	40.75940535	40.882996	67.3272093	43.48768199
Run1- $\delta/2$;Run2+ $\delta/2$	1.651820205	1.063710742	40.75940535	40.882996	67.3272093	43.48768199
Run2+ δ	1.651820205	1.063710742	40.75940535	40.882996	67.3272093	43.48768199

Fig.1 : Tableau des \log_{10} des rapports de GMT corrigés suivant les trois types de correction.

Démonstration sur un cas particulier de l'invariabilité du rapport des $\log_{10}(GMT)$.

Soit N l'effectif total de l'échantillon étudié et n_i l'effectif du bridge pour le jour i . Naturellement, $n_{35} = 0$. Fixons-nous un jour j , et comparons les expressions des $\log_{10}(GMT)$ au jour j pour chacune des trois correction :

1. Correction Run1- δ :

Soit Y_i la valeur du $\log_{10}(HI)$ associée au sujet i pour le jour j fixé. Soit X_k le titre du k -ème sujet du bridge ayant été réalisé au jour j .

D'où on obtient comme correction :

$$Y'_i = \begin{cases} Y_i - \delta \forall i \in \{1, \dots, N\} \text{ si } j = 0 \text{ ou } 21 \\ Y_i \text{ sinon} \end{cases}$$

$$X'_k = X_k \forall k \in \{1, \dots, n_j\}$$

2. Correction Run1- $\delta/2$ Run2+ $\delta/2$:

Soit Y_i la valeur du $\log_{10}(HI)$ associée au sujet i pour le jour j fixé. Soit X_k le titre du k -ème sujet du bridge ayant été réalisé au jour j .

D'où on obtient comme correction :

$$Y'_i = \begin{cases} Y_i - \delta/2 \forall i \in \{1, \dots, N\} \text{ si } j = 0 \text{ ou } 21 \\ Y_i + \delta/2 \forall i \in \{1, \dots, N\} \text{ si } j = 35 \end{cases}$$

$$X'_k = X_k + \delta/2 \forall k \in \{1, \dots, n_j\}$$

3. Correction Run2+ δ :

Soit Y_i la valeur du $\log_{10}(HI)$ associée au sujet i pour le jour j fixé. Soit X_k le titre du k -ème sujet du bridge (Run2) ayant été réalisé au jour j .

D'où on obtient comme correction :

$$Y'_i = \begin{cases} Y_i + \delta \forall i \in \{1, \dots, N\} \text{ si } j = 35 \\ Y_i \forall i \in \{1, \dots, N\} \text{ si } j = 0 \text{ ou } 21 \end{cases}$$

$$X'_k = X_k + \delta \forall k \in \{1, \dots, n_j\}$$

Vérifions donc la différence, pour un jour fixé, entre les moyennes corrigées de deux façons différentes:

On suppose que l'échantillon soit classé de telle sorte que les éléments du Run1 appartenant au bridging du jour j soient en premières positions.

Si $j=0$ ou 21, Correction3 - Correction2 :

$$\log_{10}(\text{GMT})_j = \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N Y'_i + \frac{1}{n_j} \sum_{k=1}^{n_j} \left(\frac{1}{2} \cdot (Y'_k + X'_k) \right) \right]$$

D'où on obtient que pour la correction3 on a :

$$\begin{aligned} \log_{10}(\text{GMT})_{j,3} &= \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N Y_i + \frac{1}{n_j} \sum_{k=1}^{n_j} \left(\frac{1}{2} \cdot (Y_k + X_k + \delta) \right) \right] \\ &= \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N Y_i + \frac{1}{2n_j} \sum_{k=1}^{n_j} (Y_k + X_k) + \frac{\delta}{2} \right] \\ &= \frac{\delta}{4} + \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N Y_i + \frac{1}{2n_j} \sum_{k=1}^{n_j} (Y_k + X_k) \right] \end{aligned}$$

Et pour la Correction2 on a :

$$\begin{aligned} \log_{10}(\text{GMT})_{j,2} &= \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N (Y_i - \frac{\delta}{2}) + \frac{1}{n_j} \sum_{k=1}^{n_j} \left(\frac{1}{2} \cdot (Y_k - \frac{\delta}{2} + X_k + \frac{\delta}{2}) \right) \right] \\ &= \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N (Y_i - \frac{\delta}{2}) + \frac{1}{2n_j} \sum_{k=1}^{n_j} (Y_k + X_k) \right] \\ &= \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N Y_i - \frac{\delta}{2} + \frac{1}{2n_j} \sum_{k=1}^{n_j} (Y_k + X_k) \right] \\ &= -\frac{\delta}{4} + \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N Y_i + \frac{1}{2n_j} \sum_{k=1}^{n_j} (Y_k + X_k) \right] \end{aligned}$$

Donc en faisant la différence entre $\log_{10}(\text{GMT})_3$ et $\log_{10}(\text{GMT})_2$ on obtient :

$$\begin{aligned} \log_{10}(\text{GMT})_{j,3} - \log_{10}(\text{GMT})_{j,2} &= \left(\frac{\delta}{4} + \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N Y_i + \frac{1}{2n_j} \sum_{k=1}^{n_j} (Y_k + X_k) \right] \right) \\ &\quad - \left(-\frac{\delta}{4} + \frac{1}{2} \cdot \left[\frac{1}{N-n_j} \sum_{i=n_j+1}^N Y_i + \frac{1}{2n_j} \sum_{k=1}^{n_j} (Y_k + X_k) \right] \right) \\ &= \frac{\delta}{2} \end{aligned}$$

Si $j=35$, il n'y a pas d'éléments appartenant au bridge donc on a simplement :

Pour la correction3 :

$$\begin{aligned} \log_{10}(\text{GMT})_{35,3} &= \frac{1}{N} \cdot \sum_{i=1}^N Y'_i \\ &= \frac{1}{N} \cdot \sum_{i=1}^N (Y_i + \delta) \\ &= \delta + \frac{1}{N} \cdot \sum_{i=1}^N Y_i \end{aligned}$$

Et pour la correction2 :

$$\begin{aligned} \log_{10}(\text{GMT})_{35,2} &= \frac{1}{N} \cdot \sum_{i=1}^N Y'_i \\ &= \frac{1}{N} \cdot \sum_{i=1}^N (Y_i + \frac{\delta}{2}) \\ &= \frac{\delta}{2} + \frac{1}{N} \cdot \sum_{i=1}^N Y_i \end{aligned}$$

$$\text{D'où } \log_{10}(\text{GMT})_{35,3} - \log_{10}(\text{GMT})_{35,2} = \frac{\delta}{2}$$

Donc on remarque bien que, si l'on prend l'exemple \log_{10} du rapport des GMT entre le jour 35 et 21 c'est-à-dire la différence des \log_{10} des GMT, pour la correction3 et pour la correction2 ce sera le même :

$$\begin{aligned} \log_{10}(\text{GMT})_{35,3} - \log_{10}(\text{GMT})_{21,3} &= \gamma \\ &= \log_{10}(\text{GMT})_{35,3} - \log_{10}(\text{GMT})_{21,3} + \frac{\delta}{2} - \frac{\delta}{2} \\ &= \log_{10}(\text{GMT})_{35,3} - \log_{10}(\text{GMT})_{21,3} \\ &\quad + [(\log_{10}(\text{GMT})_{21,3} - \log_{10}(\text{GMT})_{21,2}) - (\log_{10}(\text{GMT})_{35,3} - \log_{10}(\text{GMT})_{35,2})] \\ &= \log_{10}(\text{GMT})_{35,2} - \log_{10}(\text{GMT})_{21,2} \end{aligned}$$

On a donc bien démontré dans le cas particulier entre la correction3 et la correction2 que les rapports des \log_{10} des GMT corrigés restaient les mêmes. Les démonstrations pour les deux autres possibilités sont analogues à celle-ci.

• **Annexe 7: Estimation Théorique de $\text{Cov}(\overline{Y_{t,..,k,j}}, \overline{\delta})$.**

Supposons que l'on ai classé les échantillons des deux groupes de tel sorte que les n_k premiers sujets soient ceux sur lesquels ont été effectués un bridge eux même classé par ordre croissant des jours sur lesquels ont été effectués le bridge.

$$\begin{aligned} \text{Cov}(\overline{Y_{t,..,k,j}}, \overline{\delta}) &= \text{Cov}[(1/N_k) \cdot (\sum_{i=1}^{n_k} Y_{t,i,k,j} + \sum_{i=n_k+1}^{N_k} Y_{t,i,k,j}), (1/29) \cdot \sum_{q=1}^2 \sum_{l=1}^{n_q} (Y_{t',l,q,1} - Y_{t',l,q,2})] \\ &= (1/(29 \cdot N_k)) \cdot [\sum_{i=1}^{n_k} \sum_{q=1}^2 \sum_{l=1}^{n_q} \text{Cov}(Y_{t,i,k,j}, Y_{t',l,q,1}) - \sum_{i=1}^{n_k} \sum_{q=1}^2 \sum_{l=1}^{n_q} \text{Cov}(Y_{t,i,k,j}, Y_{t',l,q,2}) \\ &\quad + \sum_{i=n_k+1}^{N_k} \sum_{q=1}^2 \sum_{l=1}^{n_q} \text{Cov}(Y_{t,i,k,j}, Y_{t',l,q,1}) - \sum_{i=n_k+1}^{N_k} \sum_{q=1}^2 \sum_{l=1}^{n_q} \text{Cov}(Y_{t,i,k,j}, Y_{t',l,q,2})] \end{aligned}$$

avec $Y_{t',l,q,j}$ les observations du bridge au temps $t' \in \{0, 21\}$ pour le sujet l du groupe q au run j .

On pose comme étant fixes les effets liés au jour et au groupe. En effet, on veut pouvoir à la fin comparer l'évolution au cours du temps pour chacun des deux groupes. Ces deux effets n'entrent donc pas dans les composantes des différentes covariances puisque ne possédant pas de variabilité. On va donc retrouver des covariances qui vont s'écrire en terme de variabilité inter-sujet, inter-run et résiduelle. Ces résultats ont été testé à l'aide d'une simulation SAS d'un plan complet (voir le code SAS en Annexe partie SAS), nous avons ainsi fait l'hypothèse, que je n'ai pas pu vérifier par manque de temps, que ces résultats restaient vrais dans notre plan incomplet.

- Si $t=0$, $Y_{0,i,k,j}$ sera du run l ie $j=1$ d'où:

$$\begin{aligned} \text{Cov}(Y_{0,i,k,j}, Y_{t',l,q,1}) &= \text{Cov}(Y_{0,i,k,1}, Y_{t',l,q,1}) = \begin{cases} \text{Var}(Y_{0,i,k,1}) & \text{si } t' = 0, q = k \text{ et } i = l \\ \text{Cov}(Y_{0,i,k,1}, Y_{t',l,k,1}) & \text{sinon} \end{cases} \\ &= \begin{cases} \sigma_{B-Pid}^2 + \sigma_{B-Run}^2 + \sigma_{Res}^2 & \text{si } t' = 0, q = k \text{ et } i = l \\ \sigma_{B-Run}^2 & \text{si } t' = 0 \text{ et } \begin{cases} i \neq l \\ \text{ou} \\ i = l \text{ et } q \neq k \end{cases} \\ \sigma_{B-Pid}^2 + \sigma_{B-Run}^2 & \text{si } t' = 21, q = k \text{ et } i = l \\ \sigma_{B-Run}^2 & \text{si } t' = 21 \text{ et } \begin{cases} i \neq l \\ \text{ou} \\ i = l \text{ et } q \neq k \end{cases} \end{cases} \end{aligned}$$

$$\text{Cov}(Y_{0,i,k,j}, Y_{t',l,q,2}) = \text{Cov}(Y_{0,i,k,1}, Y_{t',l,q,2}) = \begin{cases} \sigma_{B-Pid}^2 & \text{si } q = k \text{ et } i = l \\ 0 & \text{sinon} \end{cases}$$

- Si $t=21$, $Y_{21,i,k,j}$ sera du run l ie $j=1$ d'où:

$$\text{Cov}(Y_{21,i,k,j}, Y_{t',l,q,1}) = \text{Cov}(Y_{21,i,k,1}, Y_{t',l,q,1}) = \begin{cases} \text{Var}(Y_{21,i,k,1}) & \text{si } t' = 21, q = k \text{ et } i = l \\ \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,1}) & \text{sinon} \end{cases}$$

$$= \begin{cases} \sigma_{B-Pid}^2 + \sigma_{B-Run}^2 + \sigma_{Res}^2 & si\ t' = 21, q = k\ et\ i = l \\ \sigma_{B-Run}^2 & si\ t' = 21\ et \begin{cases} i \neq l \\ ou \\ i = l\ et\ q \neq k \end{cases} \\ \sigma_{B-Pid}^2 + \sigma_{B-Run}^2 & si\ t' = 0, q = k\ et\ i = l \\ \sigma_{B-Run}^2 & si\ t' = 0\ et \begin{cases} i \neq l \\ ou \\ i = l\ et\ q \neq k \end{cases} \end{cases}$$

$$\text{Cov}(Y_{21,i,k,j}, Y_{t',l,q,2}) = \text{Cov}(Y_{21,i,k,1}, Y_{t',l,q,2}) = \begin{cases} \sigma_{B-Pid}^2 & si\ q = k\ et\ i = l \\ 0 & sinon \end{cases}$$

- Si t=35, $Y_{35,i,k,j}$ sera du run2 d'où:

$$\text{Cov}(Y_{35,i,k,j}, Y_{t',l,q,1}) = \text{Cov}(Y_{35,i,k,2}, Y_{t',l,q,1}) = \begin{cases} \sigma_{B-Pid}^2 & si\ q = k\ et\ i = l \\ 0 & sinon \end{cases}$$

$$\text{Cov}(Y_{35,i,k,j}, Y_{t',l,q,2}) = \text{Cov}(Y_{35,i,k,2}, Y_{t',l,q,2}) = \begin{cases} \sigma_{B-Pid}^2 + \sigma_{B-Run}^2 & si\ q = k\ et\ i = l \\ \sigma_{B-Run}^2 & sinon \end{cases}$$

D'où on obtient que :

- Pour le Jour 21 :

Remarque : Pour les éléments de la forme $\text{Cov}(Y_{21,i,k,1}, Y_{t',i,k',j})$, les deux i représentent des sujets différents, en effet il s'agit pour le premier du i-ème sujet du groupe k et pour le deuxième du i-ème sujet du groupe k'.

Soit $n_{k,d}$ l'effectif de l'échantillon extrait du bridge du groupe k au jour d, avec $k \in \{1,2\}$ et $d \in \{1,2\}$ avec $d=1$ si le bridge a été effectué sur l'échantillon de sang datant du jour 0 et $d=2$ s'il date du jour 21.

$$\begin{aligned} \text{Cov}(\overline{Y_{21,.,k,1}}, \overline{\delta}) &= \left(\frac{1}{(n_k + n_{k'}) \cdot N_k} \right) \cdot \left[\sum_{i=1}^{n_k} \sum_{l=1}^{n_{k,1}} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,1}) + \sum_{i=1}^{n_k} \sum_{l=1}^{n_{k',1}} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k',1}) \right. \\ &+ \sum_{i=1}^{n_k} \sum_{l=n_{k,1}+1}^{n_{k,1}+n_{k,2}} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,1}) + \sum_{i=1}^{n_k} \sum_{l=n_{k',1}+1}^{n_{k',1}+n_{k',2}} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k',1}) \\ &- \sum_{i=1}^{n_k} \sum_{l=1}^{n_k} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,2}) - \sum_{i=1}^{n_k} \sum_{l=1}^{n_{k'}} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k',2}) \\ &+ \sum_{i=n_k+1}^{N_k} \sum_{l=1}^{n_k} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,1}) + \sum_{i=n_k+1}^{N_k} \sum_{l=1}^{n_{k'}} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k',1}) \\ &\left. - \sum_{i=n_k+1}^{N_k} \sum_{l=1}^{n_k} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,2}) - \sum_{i=n_k+1}^{N_k} \sum_{l=1}^{n_{k'}} \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k',2}) \right] \\ &= \left(\frac{1}{(n_k + n_{k'}) \cdot N_k} \right) \cdot \left[n_{k,1} \cdot \text{Cov}(Y_{21,i,k,1}, Y_{0,i,k,1}) + n_{k,1} \cdot (n_k - 1) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{0,l,k,1}) \right. \\ &+ \min(n_k, n_{k',1}) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{0,i,k',1}) \\ &+ \min(n_k, n_{k',1}) \cdot (\max(n_k, n_{k',1}) - 1) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{0,l,k',1}) \\ &+ n_{k,2} \cdot \text{Var}(Y_{21,i,k,1}) + n_{k,2} \cdot (n_k - 1) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{21,l,k,1}) \\ &+ \min(n_k, n_{k',2}) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{21,i,k',1}) \\ &\left. + \min(n_k, n_{k',2}) \cdot (\max(n_k, n_{k',2}) - 1) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{21,l,k',1}) \right] \end{aligned}$$

$$\begin{aligned}
& - n_k \cdot \text{Cov}(Y_{21,i,k,1}, Y_{t',i,k,2}) - n_k \cdot (n_k - 1) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,2}) \\
& - \min(n_k, n_{k'}) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{t',i,k',2}) \\
& - \min(n_k, n_{k'}) \cdot (\max(n_k, n_{k',1}) - 1) \cdot \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k',2}) \\
& + (N_k - n_k) \cdot n_k \cdot \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,1}) \\
& + (N_k - n_k) \cdot n_{k'} \cdot \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k',1}) \\
& - (N_k - n_k) \cdot n_k \cdot \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k,2}) \\
& - (N_k - n_k) \cdot n_{k'} \cdot \text{Cov}(Y_{21,i,k,1}, Y_{t',l,k',2}) \\
= & \left(\frac{1}{(n_k + n_{k'}) \cdot N_k} \right) \cdot [n_{k,1} \cdot (\sigma_{B-Pid}^2 + \sigma_{B-Run}^2) + n_{k,1} \cdot (n_k - 1) \cdot \sigma_{B-Run}^2 \\
& + \min(n_k, n_{k',1}) \cdot \sigma_{B-Run}^2 \\
& + \min(n_k, n_{k',1}) \cdot (\max(n_k, n_{k',1}) - 1) \cdot \sigma_{B-Run}^2 \\
& + n_{k,2} \cdot (\sigma_{B-Pid}^2 + \sigma_{B-Run}^2 + \sigma_{Res}^2) + n_{k,2} \cdot (n_k - 1) \cdot \sigma_{B-Run}^2 \\
& + \min(n_k, n_{k',2}) \cdot \sigma_{B-Run}^2 + \min(n_k, n_{k',2}) \cdot (\max(n_k, n_{k',2}) - 1) \cdot \sigma_{B-Run}^2 \\
& - n_k \cdot \sigma_{B-Pid}^2 - n_k \cdot (n_k - 1) \times 0 \\
& - \min(n_k, n_{k'}) \times 0 - \min(n_k, n_{k'}) \cdot (\max(n_k, n_{k',1}) - 1) \times 0 \\
& + (N_k - n_k) \cdot n_k \cdot \sigma_{B-Run}^2 + (N_k - n_k) \cdot n_{k'} \cdot \sigma_{B-Run}^2 \\
& - (N_k - n_k) \cdot n_k \times 0 - (N_k - n_k) \cdot n_{k'} \times 0] \\
= & \left(\frac{1}{(n_k + n_{k'}) \cdot N_k} \right) \cdot [(n_{k,1} + n_{k,2} - n_k) \cdot \sigma_{B-Pid}^2 + (n_{k,1} + n_{k,1} \cdot (n_k - 1) \\
& + \min(n_k, n_{k',1}) + \min(n_k, n_{k',1}) \cdot (\max(n_k, n_{k',1}) - 1) + n_{k,2} \\
& + n_{k,2} \cdot (n_k - 1) + \min(n_k, n_{k',2}) + \min(n_k, n_{k',2}) \cdot (\max(n_k, n_{k',2}) - 1) \\
& + (N_k - n_k) \cdot n_k + (N_k - n_k) \cdot n_{k'}) \cdot \sigma_{B-Run}^2 + n_{k,2} \cdot \sigma_{Res}^2] \\
= & \left(\frac{1}{(n_k + n_{k'}) \cdot N_k} \right) \cdot [(n_k \cdot (n_{k,1} + n_{k,2} + (N_k - n_k)) + (N_k - n_k) \cdot n_{k'} \\
& + \min(n_k, n_{k',1}) \cdot \max(n_k, n_{k',1}) + \min(n_k, n_{k',2}) \cdot \max(n_k, n_{k',2})) \sigma_{B-Run}^2 \\
& + n_{k,2} \cdot \sigma_{Res}^2] \\
= & \left(\frac{1}{(n_k + n_{k'}) \cdot N_k} \right) \cdot [(n_k \cdot N_k + (N_k - n_k) \cdot n_{k'} + n_k \cdot n_{k',1} + n_k \cdot n_{k',2}) \cdot \sigma_{B-Run}^2 + n_{k,2} \cdot \sigma_{Res}^2] \\
= & \left(\frac{1}{(n_k + n_{k'}) \cdot N_k} \right) \cdot [((n_k + n_{k'}) \cdot N_k + n_k \cdot (-n_{k'} + n_{k',1} + n_{k',2})) \sigma_{B-Run}^2 + n_{k,2} \cdot \sigma_{Res}^2] \\
= & \sigma_{B-Run}^2 + \frac{n_{k,2}}{(n_k + n_{k'}) \cdot N_k} \cdot \sigma_{Res}^2
\end{aligned}$$

- Pour le Jour 0 :

Le calcul est identique à celui effectué ci-dessus pour le jour 21 et on aboutit ainsi au résultat suivant :

$$\text{Cov}(\overline{Y_{0,..,k,1}}, \overline{\delta}) = \sigma_{B-Run}^2 + \frac{n_{k,1}}{(n_k + n_{k'}) \cdot N_k} \cdot \sigma_{Res}^2$$

- Pour le Jour 35 :

Remarque : Pour les éléments de la forme $\text{Cov}(Y_{35,i,k,2}, Y_{t',i,k',j})$, les deux i représentent des sujets différents, en effet il s'agit pour le premier du i -ème sujet du groupe k et pour le deuxième du i -ème sujet du groupe k' .

$$\begin{aligned}
\text{Cov}(\overline{Y_{35,..,k,2}}, \overline{\delta}) &= \left(\frac{1}{(n_k+n_{k'}) \cdot N_k} \right) \cdot \left[\sum_{i=1}^{n_k} \sum_{l=1}^{n_k} \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k,1}) + \sum_{i=1}^{n_k} \sum_{l=1}^{n_{k'}} \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k',1}) \right. \\
&\quad - \sum_{i=1}^{n_k} \sum_{l=1}^{n_k} \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k,2}) - \sum_{i=1}^{n_k} \sum_{l=1}^{n_{k'}} \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k',2}) \\
&\quad + \sum_{i=n_k+1}^{N_k} \sum_{l=1}^{n_k} \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k,1}) + \sum_{i=n_k+1}^{N_k} \sum_{l=1}^{n_{k'}} \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k',1}) \\
&\quad \left. - \sum_{i=n_k+1}^{N_k} \sum_{l=1}^{n_k} \text{Cov}(Y_{35,i,k,1}, Y_{t',l,k,2}) - \sum_{i=n_k+1}^{N_k} \sum_{l=1}^{n_{k'}} \text{Cov}(Y_{35,i,k,1}, Y_{t',l,k',2}) \right] \\
&= \left(\frac{1}{(n_k+n_{k'}) \cdot N_k} \right) \cdot \left[n_k \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',i,k,1}) + n_k \cdot (n_k - 1) \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k,1}) \right. \\
&\quad + \min(n_k, n_{k'}) \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',i,k',1}) \\
&\quad + \min(n_k, n_{k'}) \cdot (\max(n_k, n_{k'}) - 1) \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k',1}) \\
&\quad - n_k \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',i,k,2}) - n_k \cdot (n_k - 1) \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k,2}) \\
&\quad - \min(n_k, n_{k'}) \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',i,k',2}) \\
&\quad - \min(n_k, n_{k'}) \cdot (\max(n_k, n_{k'}) - 1) \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k',2}) \\
&\quad + (N_k - n_k) \cdot n_k \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k,1}) \\
&\quad + (N_k - n_k) \cdot n_{k'} \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k',1}) - (N_k - n_k) \cdot n_k \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k,2}) \\
&\quad \left. - (N_k - n_k) \cdot n_{k'} \cdot \text{Cov}(Y_{35,i,k,2}, Y_{t',l,k',2}) \right] \\
&= \left(\frac{1}{(n_k+n_{k'}) \cdot N_k} \right) \cdot \left[n_k \cdot \sigma_{B-Pid}^2 + n_k \cdot (n_k - 1) \times 0 + \min(n_k, n_{k'}) \times 0 \right. \\
&\quad + \min(n_k, n_{k'}) \cdot (\max(n_k, n_{k'}) - 1) \times 0 \\
&\quad - n_k \cdot (\sigma_{B-Run}^2 + \sigma_{B-Pid}^2) - n_k \cdot (n_k - 1) \cdot \sigma_{B-Run}^2 \\
&\quad - \min(n_k, n_{k'}) \cdot \sigma_{B-Run}^2 - \min(n_k, n_{k'}) \cdot (\max(n_k, n_{k'}) - 1) \cdot \sigma_{B-Run}^2 \\
&\quad + (N_k - n_k) \cdot n_k \times 0 \\
&\quad + (N_k - n_k) \cdot n_{k'} \times 0 - (N_k - n_k) \cdot n_k \cdot \sigma_{B-Run}^2 \\
&\quad \left. - (N_k - n_k) \cdot n_{k'} \cdot \sigma_{B-Run}^2 \right] \\
&= \left(\frac{1}{(n_k+n_{k'}) \cdot N_k} \right) \cdot \left[((n_k - n_k) \cdot \sigma_{B-Pid}^2 + (-n_k - n_k \cdot (n_k - 1)) \right. \\
&\quad - \min(n_k, n_{k'}) - \min(n_k, n_{k'}) \cdot (\max(n_k, n_{k'}) - 1) \\
&\quad \left. - (N_k - n_k) \cdot n_k - (N_k - n_k) \cdot n_{k'} \cdot \sigma_{B-Run}^2 \right] \\
&= \left(\frac{1}{(n_k+n_{k'}) \cdot N_k} \right) \cdot \left[(-n_k - n_k \cdot (n_k - 1)) \right. \\
&\quad \left. - \min(n_k, n_{k'}) \cdot \max(n_k, n_{k'}) - (N_k - n_k) \cdot n_k - (N_k - n_k) \cdot n_{k'} \cdot \sigma_{B-Run}^2 \right] \\
&= \left(\frac{1}{(n_k+n_{k'}) \cdot N_k} \right) \cdot \left[(-n_k - n_k \cdot (n_k - 1)) - n_k \cdot n_{k'} - (N_k - n_k) \cdot n_k - (N_k - n_k) \cdot n_{k'} \cdot \sigma_{B-Run}^2 \right] \\
&= \left(\frac{1}{(n_k+n_{k'}) \cdot N_k} \right) \cdot \left[(-n_k \cdot n_k - n_k \cdot n_{k'} - (N_k - n_k) \cdot (n_k + n_{k'})) \cdot \sigma_{B-Run}^2 \right] \\
&= \left(\frac{1}{(n_k+n_{k'}) \cdot N_k} \right) \cdot \left[-(N_k - n_k + n_k) \cdot (n_k + n_{k'}) \cdot \sigma_{B-Run}^2 \right] \\
&= - \frac{1}{(n_k+n_{k'}) \cdot N_k} \cdot N_k \cdot (n_k + n_{k'}) \cdot \sigma_{B-Run}^2 \\
&= - \sigma_{B-Run}^2
\end{aligned}$$

• **Annexe 9: Etude de la variance du log₁₀ du rapport des GMT corrigés entre jour 35 et 21**

Soit $Y'_{t,i,k}$ la variable réponse log₁₀(HI) après correction au temps t, pour le sujet i du groupe k, avec $t \in \{0, 21, 35\}$, $k \in \{1, 2\}$, $i \in \{1, \dots, N_k\}$ où N_k est l'effectif total associé au groupe k.

et $Y_{t,i,k,j}$ la variable réponse log₁₀(HI) avant correction au temps t, pour le sujet i du groupe k, au run j
avec $t \in \{0, 21, 35\}$, $k \in \{1, 2\}$, $i \in \{1, \dots, N_k\}$, $j \in \{1, 2\}$.

Nous allons nous intéresser tout particulièrement à la variance de la différence des moyennes des log₁₀(HI) entre le jour 35 et le jour 21, c'est-à-dire à la variance du log₁₀ du rapport des GMT entre le jour 35 et le jour 21, puisque c'est à ce moment que se produit le changement de Run. Cependant, il est important de noter qu'en faisant cette étude sur les variables corrigées à l'aide du $\bar{\delta}$. obtenue comme étant l'écart moyen entre le run1 et le run2, une partie de l'écart et de la variabilité entre ces deux journées sera encore dû à l'effet run. Si l'on voulait bien faire et réellement comparer que ces deux jours, il aurait fallu recalculer un delta représentant cette fois-ci l'effet run obtenu uniquement à l'aide des données du bridge appartenant au jour 21. Cela aurait permis de réaliser une correction exacte et ainsi pouvoir évaluer la variance exacte du rapport des GMT corrigées, mais corrigées d'une autre façon. Cependant, les rapports entre les titres obtenus au bridge et les titres initiaux correspondants, restent presque constants quelque soit la valeur de base du titre donc on peut supposer que si l'on recalculait ce delta on ne serait pas loin de retrouver la même valeur (Voir Annexe1). Si l'on se retrouvait en face d'une situation où ce rapport croît ou décroît lorsque les valeurs du Run1 augmente ou diminue alors il faudrait se poser la question et certainement décider de recalculer ce delta pour ce calcul. De plus, la correction avec un $\bar{\delta}$. calculé uniquement sur la base des données du bridge du 21eme jour aurait du sens si l'on ne voulait que comparer les données du jour 21 et 35, or comme dans cette étude le but principal n'est pas de rectifier l'effet run entre le jour 21 et 35, mais bien l'effet run entre le Run1 et le Run2 en général, on va garder ce $\bar{\delta}$. et accepter la légère erreur d'interprétation que cela apporte sur l'estimation de la variance de ce rapport.

$$\begin{aligned}
 \text{Var}(\overline{Y'_{35,..,k}} - \overline{Y'_{21,..,k}}) &= \text{Var}(\overline{Y'_{35,..,k}}) + \text{Var}(\overline{Y'_{21,..,k}}) - 2.\text{Cov}(\overline{Y'_{35,..,k}}, \overline{Y'_{21,..,k}}) \\
 &= \text{Var}(\overline{Y_{35,..,k,2} + \bar{\delta}/2}) + \text{Var}(\overline{Y_{21,..,k,1} - \bar{\delta}/2}) - 2.\text{Cov}(\overline{Y_{35,..,k,2} + \bar{\delta}/2}, \overline{Y_{21,..,k,1} - \bar{\delta}/2}) \\
 &= \text{Var}(\overline{Y_{35,..,k,2}}) + \text{Var}(\overline{\bar{\delta}/2}) + 2.\text{Cov}(\overline{Y_{35,..,k,2}}, \overline{\bar{\delta}/2}) + \text{Var}(\overline{Y_{21,..,k,1}}) + \text{Var}(\overline{\bar{\delta}/2}) \\
 &\quad - 2.\text{Cov}(\overline{Y_{21,..,k,1}}, \overline{\bar{\delta}/2}) - 2.[\text{Cov}(\overline{Y_{35,..,k,2}}, \overline{Y_{21,..,k,1}}) \\
 &\quad - \text{Cov}(\overline{Y_{35,..,k,2}}, \overline{\bar{\delta}/2}) + \text{Cov}(\overline{Y_{21,..,k,1}}, \overline{\bar{\delta}/2}) - \text{Cov}(\overline{\bar{\delta}/2}, \overline{\bar{\delta}/2})] \\
 &= 4.\text{Var}(\overline{\bar{\delta}/2}) + \text{Var}(\overline{Y_{35,..,k,2}}) + \text{Var}(\overline{Y_{21,..,k,1}}) + 4.\text{Cov}(\overline{Y_{35,..,k,2}}, \overline{\bar{\delta}/2}) \\
 &\quad - 4.\text{Cov}(\overline{Y_{21,..,k,1}}, \overline{\bar{\delta}/2}) - 2.\text{Cov}(\overline{Y_{35,..,k,2}}, \overline{Y_{21,..,k,1}}) \\
 &= \text{Var}(\overline{\bar{\delta}}) + \text{Var}(\overline{Y_{35,..,k,2}}) + \text{Var}(\overline{Y_{21,..,k,1}}) \\
 &\quad + 2.[\text{Cov}(\overline{Y_{35,..,k,2}}, \overline{\bar{\delta}}) - \text{Cov}(\overline{Y_{21,..,k,1}}, \overline{\bar{\delta}}) - \text{Cov}(\overline{Y_{35,..,k,2}}, \overline{Y_{21,..,k,1}})]
 \end{aligned}$$

Il est simple d'estimer de façon empirique $\text{Var}(\overline{\bar{\delta}})$, $\text{Var}(\overline{Y_{35,..,k,2}})$ et $\text{Var}(\overline{Y_{21,..,k,1}})$. Ainsi, pour calculer $\text{Var}(\overline{\bar{\delta}})$, commençons par calculer la variance des δ_i . A l'aide de la proc Means de SAS, on obtient comme résultat (voir Annexe10):

$$\text{Var}(\delta_i) = 0.01514$$

Ceci aurait naturellement pu être calculé à la main grâce à la formule classique pour l'estimation d'une variance à savoir :

$$\frac{1}{28} \sum_{i=1}^{29} (\delta_i - \bar{\delta})^2 = \frac{1}{28} \sum_{i=1}^{29} (\delta_i - 0.1354)^2 = 0.01514$$

D'où on en déduit $\text{Var}(\bar{\delta}) = \frac{1}{29} \text{Var}(\delta_i)$. En effet, comme les $\delta_i \stackrel{iid}{\sim}$ Normale alors la moyenne $\bar{\delta}$ suit aussi une loi normale de variance la variance des δ_i divisée par 29. On trouve ainsi que:

$$\text{Var}(\bar{\delta}) = 0.000522.$$

Ensuite, pour $\text{Var}(\overline{Y_{35,..,k,2}})$, le principe reste le même. Ces variances ont été estimées à l'aide de la proc Mixed de SAS (voir Annexe10). On obtient ainsi :

$$\text{Var}(\overline{Y_{35,..,1,2}}) = 0.002721 ; \text{Var}(\overline{Y_{35,..,2,2}}) = 0.002543$$

Enfin, pour le jour 21 on obtient :

$$\text{Var}(\overline{Y_{21,..,1,1}}) = 0.003873 ; \text{Var}(\overline{Y_{21,..,2,1}}) = 0.003651$$

Ensuite, $\text{Cov}(\overline{Y_{35,..,k,2}}, \bar{\delta})$ et $\text{Cov}(\overline{Y_{21,..,k,1}}, \bar{\delta})$ ont déjà été estimés (voir Annexe7). Il ne reste donc plus qu'à calculer $\text{Cov}(\overline{Y_{35,..,k,2}}, \overline{Y_{21,..,k,1}})$:

$$\begin{aligned} \text{Cov}(\overline{Y_{35,..,k,2}}, \overline{Y_{21,..,k,1}}) &= \text{Cov}\left[\frac{1}{N_k} \cdot \sum_{i=1}^{N_k} Y_{35,i,k,2}, \frac{1}{N_k} \cdot \sum_{j=1}^{N_k} Y_{21,j,k,1}\right] \\ &= \left(\frac{1}{N_k}\right)^2 \cdot \sum_{i=1}^{N_k} \text{Cov}(Y_{35,i,k,2}, Y_{21,i,k,1}), \text{ par indep des sujets.} \\ &= \left(\frac{1}{N_k}\right)^2 \cdot \sum_{i=1}^{N_k} \sigma_{B-Pid}^2 \\ &= \left(\frac{1}{N_k}\right) \cdot \sigma_{B-Pid}^2 \end{aligned}$$

D'où on obtient finalement que :

$$\begin{aligned} \text{Var}(\overline{Y'_{35,..,k}} - \overline{Y'_{21,..,k}}) &= \text{Var}(\bar{\delta}) + \text{Var}(\overline{Y_{35,..,k,2}}) + \text{Var}(\overline{Y_{21,..,k,1}}) \\ &\quad + 2 \cdot [\text{Cov}(\overline{Y_{35,..,k,2}}, \bar{\delta}) - \text{Cov}(\overline{Y_{21,..,k,1}}, \bar{\delta}) - \text{Cov}(\overline{Y_{35,..,k,2}}, \overline{Y_{21,..,k,1}})] \\ &= \text{Var}(\bar{\delta}) + \text{Var}(\overline{Y_{35,..,k,2}}) + \text{Var}(\overline{Y_{21,..,k,1}}) \\ &\quad + 2 \cdot \left(-\sigma_{B-Run}^2 - \left(\sigma_{B-Run}^2 + \frac{n_{k,2}}{(n_k+n_{k'}) \cdot N_k} \cdot \sigma_{Res}^2 \right) - \left(\frac{1}{N_k}\right) \cdot \sigma_{B-Pid}^2 \right) \\ &= \text{Var}(\bar{\delta}) + \text{Var}(\overline{Y_{35,..,k,2}}) + \text{Var}(\overline{Y_{21,..,k,1}}) \\ &\quad + 2 \cdot \left(-2 \cdot \sigma_{B-Run}^2 - \frac{n_{k,2}}{(n_k+n_{k'}) \cdot N_k} \cdot \sigma_{Res}^2 - \left(\frac{1}{N_k}\right) \cdot \sigma_{B-Pid}^2 \right) \end{aligned}$$

Donc pour le groupe 1 on obtient:

$$\begin{aligned} \text{Var}(\overline{Y'_{35,..,1}} - \overline{Y'_{21,..,1}}) &= \text{Var}(\bar{\delta}) + \text{Var}(\overline{Y_{35,..,1,2}}) + \text{Var}(\overline{Y_{21,..,1,1}}) \\ &\quad + 2 \cdot \left(-2 \cdot \sigma_{B-Run}^2 - \frac{6}{928} \cdot \sigma_{Res}^2 - \left(\frac{1}{64}\right) \cdot \sigma_{B-Pid}^2 \right) \end{aligned}$$

Et pour le groupe 2 on obtient:

$$\begin{aligned} \text{Var}(\overline{Y'_{35,..,2}} - \overline{Y'_{21,..,2}}) &= \text{Var}(\bar{\delta}) + \text{Var}(\overline{Y_{35,..,2,2}}) + \text{Var}(\overline{Y_{21,..,2,1}}) \\ &\quad + 2 \cdot \left(-2 \cdot \sigma_{B-Run}^2 - \frac{4}{957} \cdot \sigma_{Res}^2 - \left(\frac{1}{66}\right) \cdot \sigma_{B-Pid}^2 \right) \end{aligned}$$

Application Numérique (Voir Annexe 8):

Pour le groupe 1, on obtient :

$$\begin{aligned}\text{Var}(\overline{Y'_{35,..,1}} - \overline{Y'_{21,..,1}}) &= 0.000522 + 0.002721 + 0.003873 + 2.(-2 \cdot \sigma_{B-Run}^2 - \frac{6}{928} \cdot \sigma_{Res}^2 - \frac{1}{64} \cdot \sigma_{B-Pid}^2) \\ &= 0.000522 + 0.002721 + 0.003873 + 2.(-2 \times 0.000976 - \frac{6}{928} \times 0.00757 - \frac{1}{64} \times 0.09624) \\ &= 0.000107\end{aligned}$$

Pour le groupe 2, on obtient :

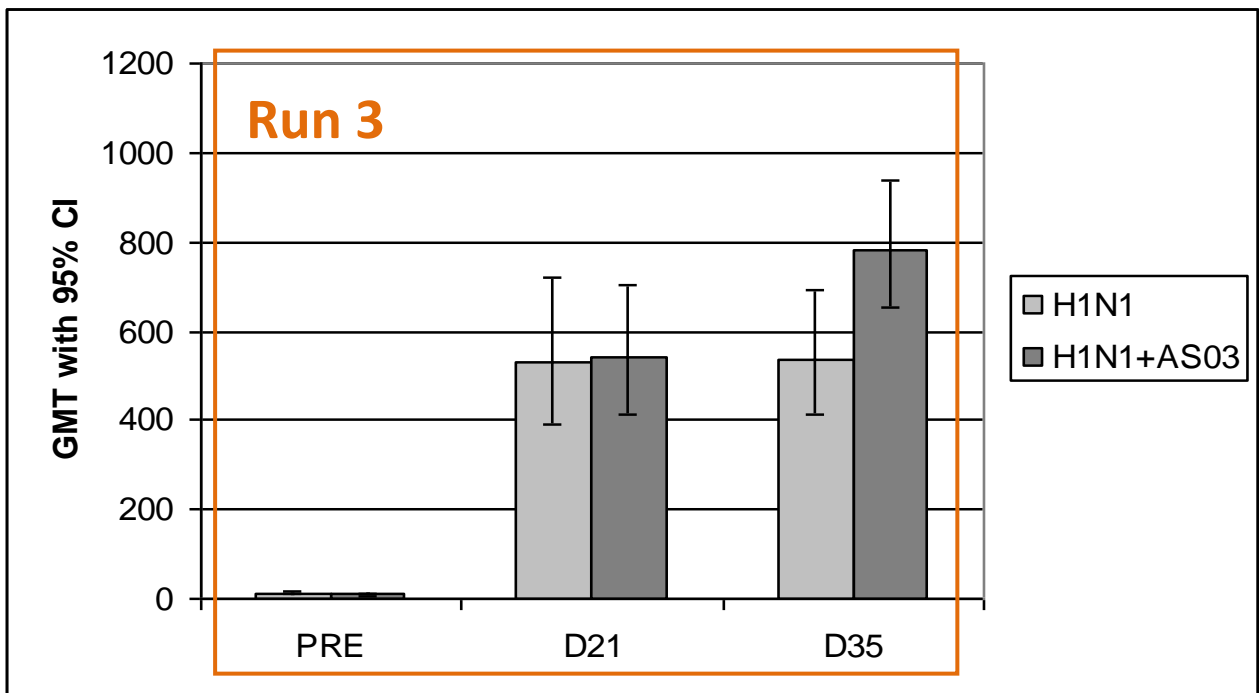
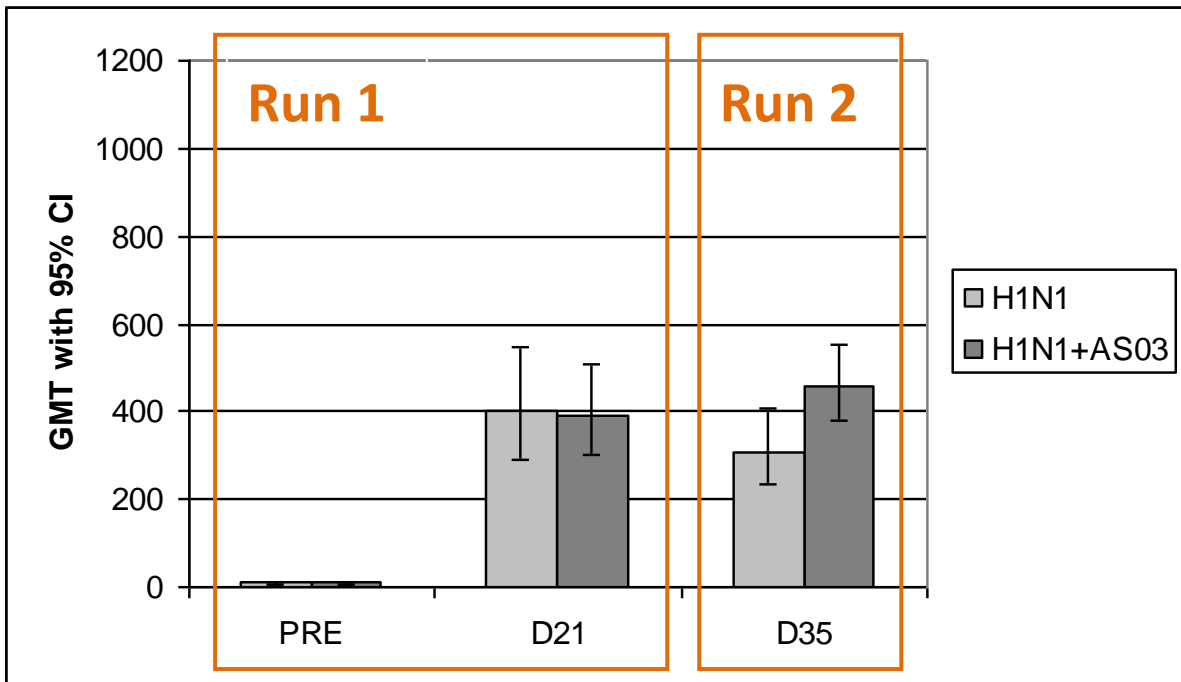
$$\begin{aligned}\text{Var}(\overline{Y'_{35,..,2}} - \overline{Y'_{21,..,2}}) &= 0.000522 + 0.002543 + 0.003651 + 2.(-2 \cdot \sigma_{B-Run}^2 - \frac{4}{957} \cdot \sigma_{Res}^2 - (\frac{1}{66}) \cdot \sigma_{B-Pid}^2) \\ &= 0.000522 + 0.002543 + 0.003651 + 2.(-2 \times 0.000976 - \frac{4}{957} \times 0.00757 - (\frac{1}{66}) \times 0.09624) \\ &= -0.000168 !!!\end{aligned}$$

On atteint ici les limites de la méthode utilisée. En effet, les différences des titres pour le groupe 2 entre le jour 35 et le jour 21 sont très faibles et possèdent donc une variance trop faible par rapport à la covariance qui les relie. Il faudrait donc étudier la possibilité d'estimer ces valeurs autrement mais par manque de temps, il ne m'a pas été possible de faire cela. La deuxième solution est de supposer l'indépendance entre l'écart inter-run moyen et les moyennes non corrigées. Ainsi la covariance entre ces termes seraient supposés nulle et on obtiendrait :

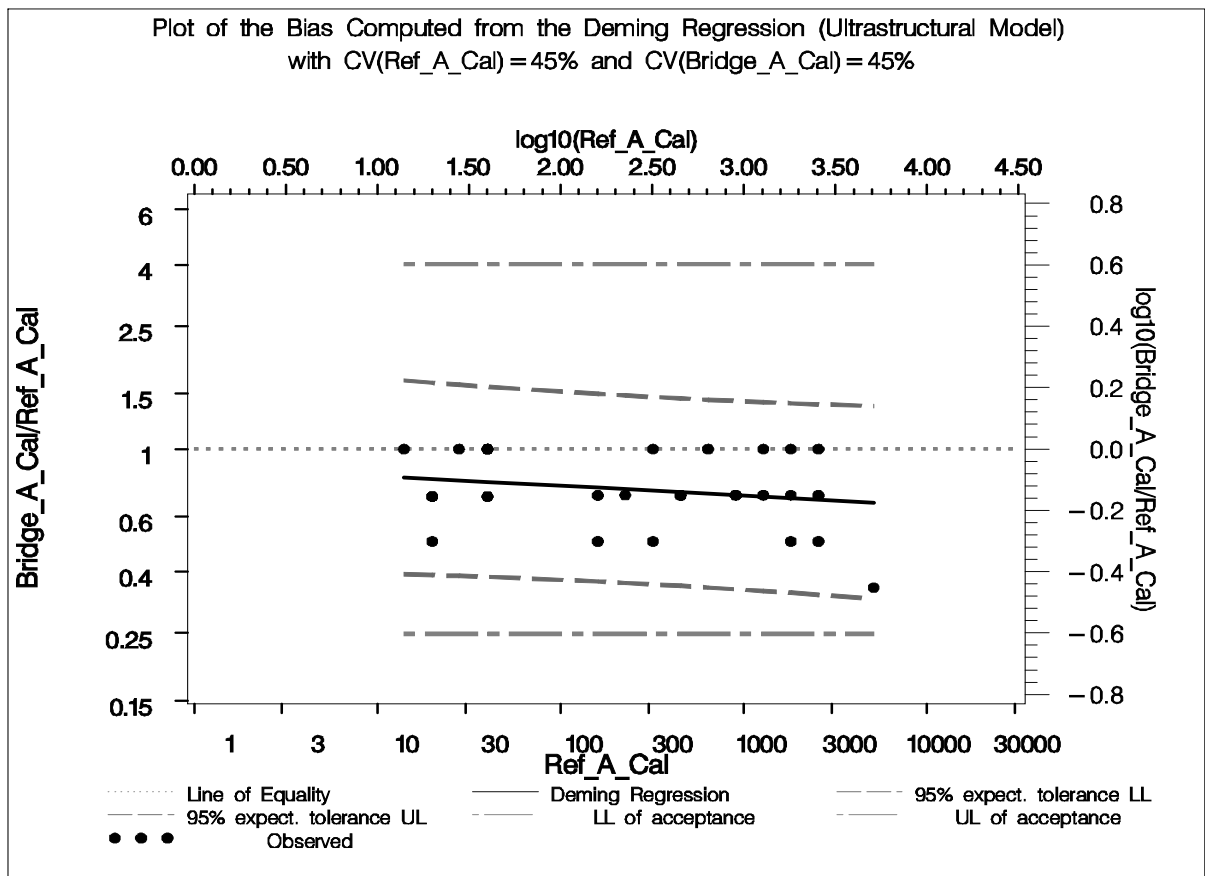
$$\begin{aligned}\text{Var}(\overline{Y'_{35,..,k}} - \overline{Y'_{21,..,k}}) &= \text{Var}(\overline{\delta.}) + \text{Var}(\overline{Y_{35,..,k,2}}) + \text{Var}(\overline{Y_{21,..,k,1}}) \\ &\quad + 2.[\text{Cov}(\overline{Y_{35,..,k,2}}, \overline{\delta.}) - \text{Cov}(\overline{Y_{21,..,k,1}}, \overline{\delta.}) - \text{Cov}(\overline{Y_{35,..,k,2}}, \overline{Y_{21,..,k,1}})] \\ &= \text{Var}(\overline{\delta.}) + \text{Var}(\overline{Y_{35,..,k,2}}) + \text{Var}(\overline{Y_{21,..,k,1}}) - 2 \cdot \text{Cov}(\overline{Y_{35,..,k,2}}, \overline{Y_{21,..,k,1}}) \\ &\geq 0 \text{ pour les deux groupes.}\end{aligned}$$

Annexes

• Annexe1 : Résultats Préliminaires

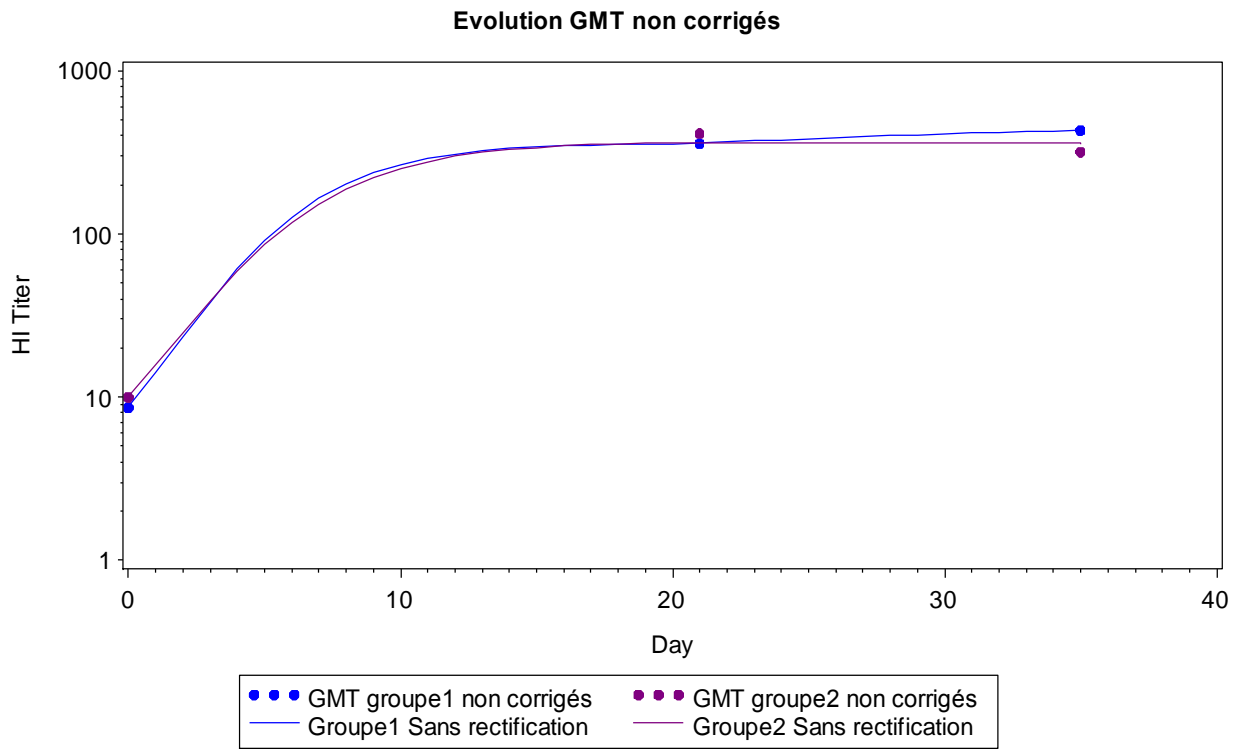


Représentation des rapports entre Run1 et Run2 des données du bridge.



Ratio Moyen : 0.73

•Annexe 2 : Exemple de représentation graphique, l'évolution des GMT non corrigés



• Annexe 3 : Tableaux de Corrélation des deux modèles utilisés

Coefficients de corrélation de Pearson, N = 384 Proba > r sous H0: Rho=0			
	run	GROUP_NB	day
run	1.00000	0.01144 0.8231	0.80166 <.0001
GROUP_NB Group number (effective)	0.01144 0.8231	1.00000	0.01306 0.7986
day	0.80166 <.0001	0.01306 0.7986	1.00000

Fig1 : Tableau de corrélation des effets du modèle (1) associé aux données du Run1 et 2 hors bridge.

Type 3 Tests des effets fixes				
Effet	DDL Num.	DDL Res.	Valeur F	Pr > F
GROUP_NB	1	250	0.00	0.9721
day	2	250	715.89	<.0001
GROUP_NB*day	2	250	11.46	<.0001

Fig2 : Résultat sur la significativité des effets du modèle (1)

Coefficients de corrélation de Pearson, N = 58 Proba > r sous H0: Rho=0			
	run	group_nb	day
run	1.00000	0.00000 1.0000	0.00000 1.0000
group_nb	0.00000 1.0000	1.00000	0.06356 0.6355
day	0.00000 1.0000	0.06356 0.6355	1.00000

Fig3 : Tableau de corrélation des effets du modèle (2) associé aux données du bridge.

Type 3 Tests des effets fixes				
Effet	DDL Num.	DDL Res.	Valeur F	Pr > F
group_nb	1	28	0.15	0.6975
day	1	28	100.21	<.0001
group_nb*day	1	28	3.59	0.0686
run	1	28	35.13	<.0001

Fig4 : Première étape sélection du modèle (2).

Type 3 Tests des effets fixes				
Effet	DDL Num.	DDL Res.	Valeur F	Pr > F
group_nb	1	28	1.51	0.2300
day	1	28	88.98	<.0001
run	1	28	35.13	<.0001

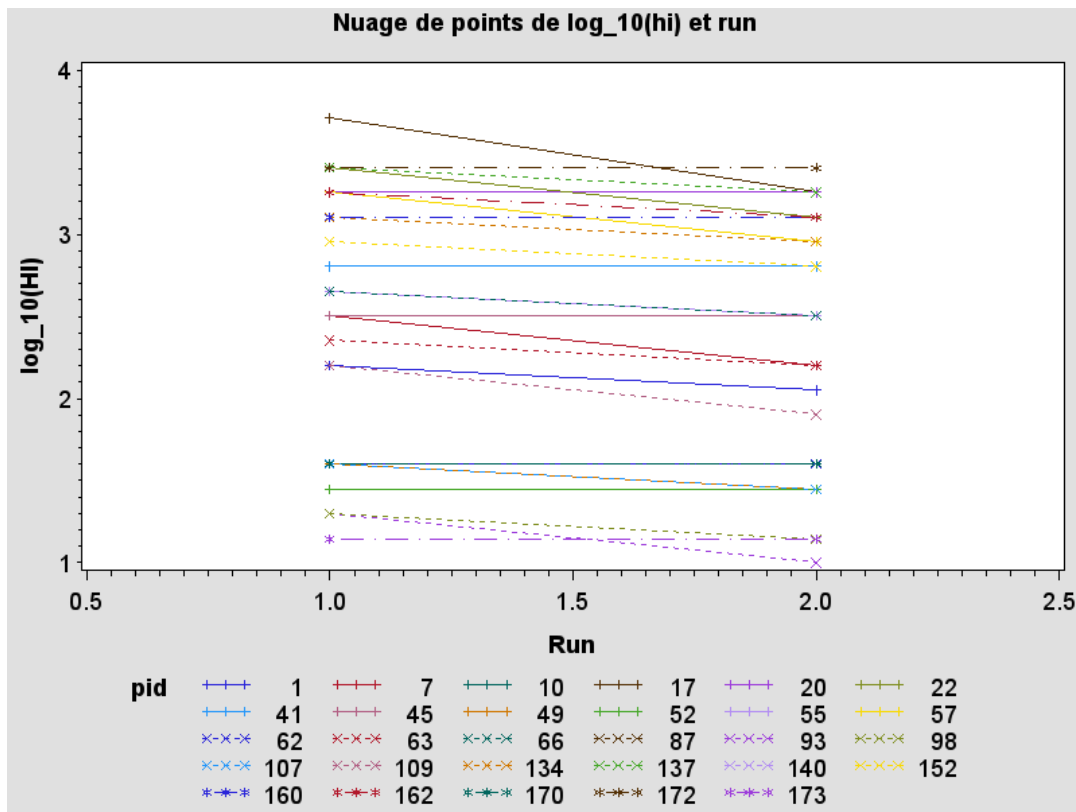
Fig5 : Deuxième étape sélection du modèle (2).

Type 3 Tests des effets fixes				
Effet	DDL Num.	DDL Res.	Valeur F	Pr > F
day	1	28	89.15	<.0001
run	1	28	35.13	<.0001

Fig6 : Troisième étape sélection du modèle (2).

Les résultats restent naturellement identiques en faisant la sélection pour l'effet run en aléatoire.

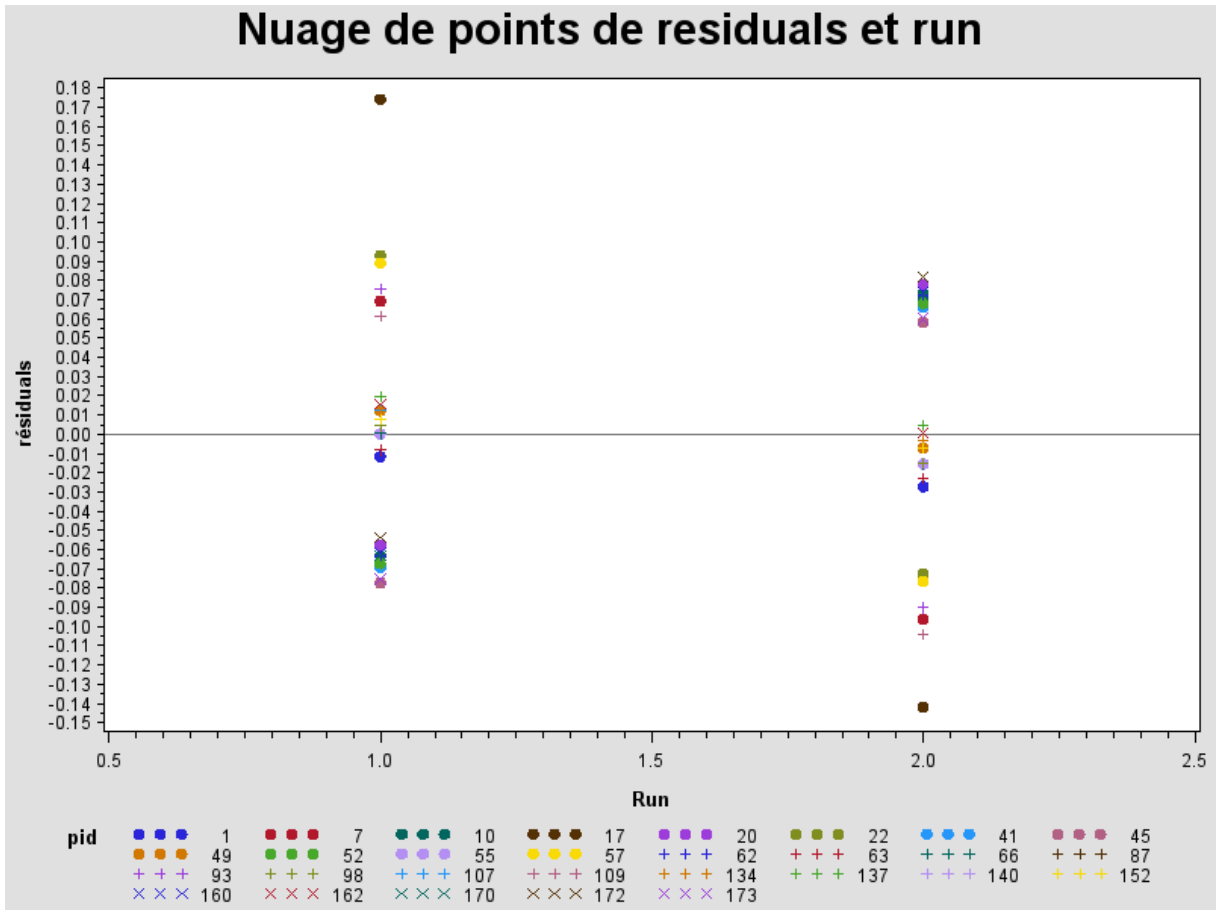
• **Annexe 4 : Exemple de vérification des hypothèses sur les résidus**



Tests de tendance centrale : $\mu_0=0$				
Test	Statistique		p Value	
t de Student	t	0	Pr > t	1.0000
Signe	M	0	Pr >= M	1.0000
Rang signé	S	-6.5	Pr >= S	0.9604

Tests de normalité				
Test	Statistique		p Value	
Shapiro-Wilk	W	0.965356	Pr < W	0.0962
Kolmogorov-Smirnov	D	0.110319	Pr > D	0.0787
Cramer-von Mises	W-Sq	0.150343	Pr > W-Sq	0.0232
Anderson-Darling	A-Sq	0.890317	Pr > A-Sq	0.0223

Bartlett's Test for Homogeneity of Resid Variance			
Source	DDL	Khi-2	Pr > Khi-2
run	1	0.0210	0.8847



• Annexe 5 : Tableaux des GMT estimés et de l'écart inter-run

Moyennes des moindres carrés										
Effet	Group number	day	Valeur estimée	Erreur type	DDL	Valeur du test t	Pr > t	Alpha	Inférieur	Supérieur
GROUP_NB *day	1	0	0.9364	0.05341	128	17.53	<.0001	0.05	0.8307	1.0421
GROUP_NB *day	1	21	2.5551	0.06223	128	41.06	<.0001	0.05	2.4319	2.6782
GROUP_NB *day	1	35	2.6374	0.05216	128	50.56	<.0001	0.05	2.5341	2.7406
GROUP_NB *day	2	0	1.0000	0.05260	128	19.01	<.0001	0.05	0.8959	1.1041
GROUP_NB *day	2	21	2.6170	0.06042	128	43.31	<.0001	0.05	2.4974	2.7365
GROUP_NB *day	2	35	2.5052	0.05043	128	49.68	<.0001	0.05	2.4054	2.6050

Fig1 : Sortie SAS de la proc mixed associée au modèle (1) portant sur les données hors bridge pour l'estimation des GMT non corrigés.

Différences des moyennes des moindres carrés							
Effet	run	_run	Valeur estimée	Erreur type	DDL	Valeur du test t	Pr > t
run	1	2	0.1354	0.02285	28	5.93	<.0001

Fig2 : Sortie SAS de la proc mixed associée au modèle (2) portant sur les données du bridge pour l'estimation de l'effet inter-run.

Moyennes des moindres carrés										
Effet	Group number	day	Valeur estimée	Erreur type	DDL	Valeur du test t	Pr > t	Alph a	Inférieur	Supérieur
GROUP_NB *day	1	0	0.8769	0.05329	128	16.46	<.0001	0.05	0.7714	0.9823
GROUP_NB *day	1	21	2.4871	0.06206	128	40.07	<.0001	0.05	2.3643	2.6099
GROUP_NB *day	1	35	2.7051	0.05219	128	51.84	<.0001	0.05	2.6018	2.8083
GROUP_NB *day	2	0	0.9345	0.05247	128	17.81	<.0001	0.05	0.8307	1.0383
GROUP_NB *day	2	21	2.5461	0.06024	128	42.26	<.0001	0.05	2.4269	2.6653
GROUP_NB *day	2	35	2.5729	0.05044	128	51.01	<.0001	0.05	2.4731	2.6727

Fig3 : Sortie SAS de la proc mixed associée au modèle (2) portant sur les données corrigées pour l'estimation des GMT corrigés (avec IC pour correction fixe).

• Annexe 8 : Tableaux des variabilités inter-run, inter-pid et résiduelle

Valeurs estimées des paramètres de covariance		
Param de cov	Sujet	Valeur estimée
run		0.000976
day*run		0.003439
CS	PID	0.09721
Residual		0.09497

Fig1 : Tableau pour la variabilité inter-run

Valeurs estimées des paramètres de covariance		
Param de cov	Sujet	Valeur estimée
CS	PID	0.09624
Residual		0.09410

Fig2 : Tableau pour la variabilité inter-pid
(Obtenu à partir du modèle (1) sur l'ensemble des données)

Valeurs estimées des paramètres de covariance		
Param de cov	Sujet	Valeur estimée
CS	pid	0.1402
Residual		0.007570

Fig3 : Tableau pour la variabilité résiduelle

• Annexe 10 : Résultats de variance

Variable d'analyse : d_run
Variance
0.0151403

Fig.1 : Variance des écarts inter-run

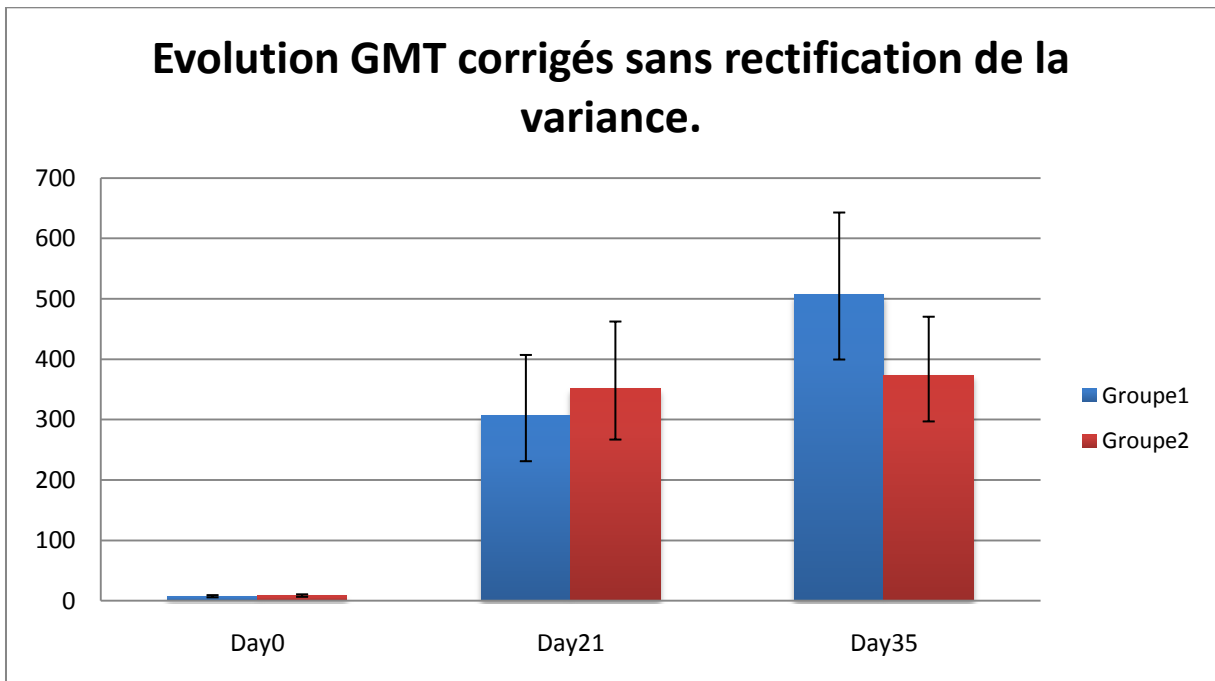
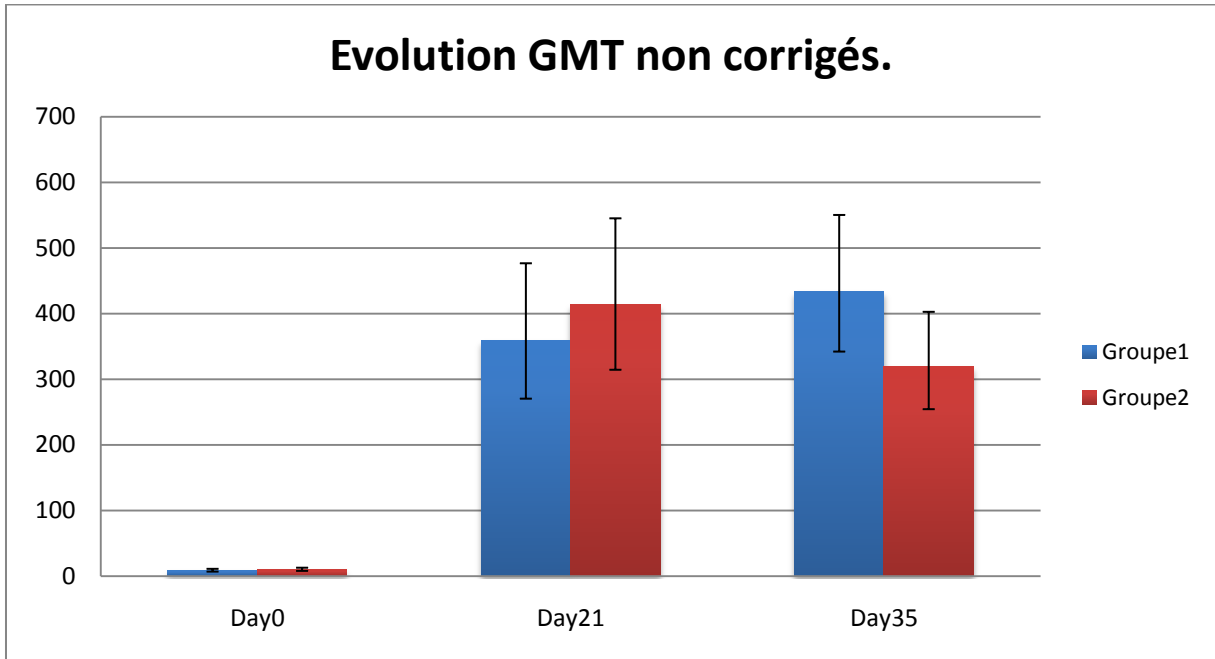
Effet	Group number	day	Value estimée	DDL	Cov1	Cov2	Cov3	Cov4	Cov5	Cov6
GROUP_NB *day	1	0	0.9364	128	0.002853	0.001100	0.000930	-434E-21	-434E-21	-434E-21
GROUP_NB *day	1	21	2.5551	128	0.001100	0.003873	0.002770	-379E-21	-217E-21	-434E-21
GROUP_NB *day	1	35	2.6374	128	0.000930	0.002770	0.002721	-434E-21	-434E-21	-434E-21
GROUP_NB *day	2	0	1.0000	128	-434E-21	-379E-21	-434E-21	0.002766	0.001066	0.000902
GROUP_NB *day	2	21	2.6170	128	-434E-21	-217E-21	-434E-21	0.001066	0.003651	0.002613
GROUP_NB *day	2	35	2.5052	128	-434E-21	-434E-21	-434E-21	0.000902	0.002613	0.002543

Fig.2 : Tableau des variances et covariances des log(GMT)

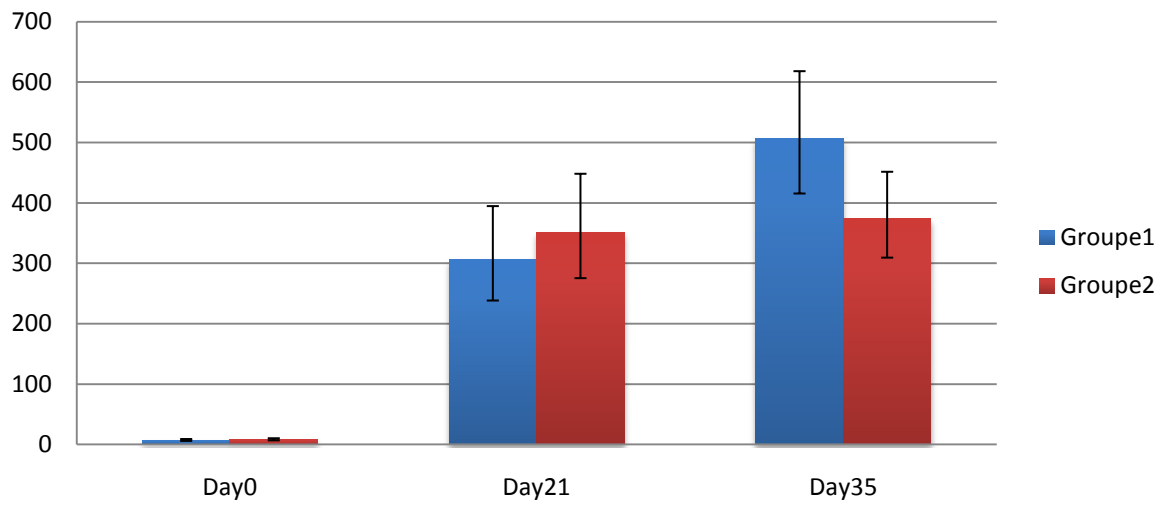
• Annexe 11 : Tableau des résultats pour IC

Obs	Effect	GRO UP_N B	day	gmt	ll	ul	covariance	variance_t ot	rho	d_f	q
1	GROUP_ NB*day	1	0	7.531	6.120	9.268	0.001000472	.001982836	0.81979	38.5160	2.02350
2	GROUP_ NB*day	1	21	306.977	238.607	394.938	0.001024944	.002978175	0.72083	53.7457	2.00510
3	GROUP_ NB*day	1	35	507.071	415.809	618.363	-.000976000	.001875662	-0.81885	80.7691	1.98977
4	GROUP_ NB*day	2	0	8.600	7.015	10.544	0.000987865	.001908995	0.82201	37.5687	2.02516
5	GROUP_ NB*day	2	21	351.606	275.647	448.496	0.001007641	.002773617	0.72987	51.5431	2.00707
6	GROUP_ NB*day	2	35	374.007	309.618	451.787	-.000976000	.001697780	-0.84701	76.6652	1.99139

• Annexe 12 : Graphiques finaux



Evolution GMT corrigés avec rectification de la variance



• Annexe 13 : Une partie du code SAS utilisée.

```
/******Code SAS pour l'analyse fréquentiste******/
```

```
options nodate nonumber nocenter;
```

```
*libname as021 '\\Rixsffs05\biometric\Biostat Sub-  
Teams\FluStat\Science\Immuno\Kinetic\Bridging';  
libname as021 'C:\Users\Darkflow\Documents\cours\Master Stat\Stage M1  
Statistique\Bridging';
```

```
data bridge;  
  input pid group_nb day hi run;  
  cards;  
  *Confidentiel*  
  ;  
run;
```

```
data one;  
  set as021.h1_021_f;  
  l_hi=log10(hi);  
run;
```

```
proc sort data=one;  
  by pid day run;  
run;
```

```
data nobridge;  
  set one;  
  if day=0 and run=2 then delete;  
  if day=21 and run=2 then delete;  
run;
```

```
proc sort data=nobridge;  
  by pid day run;  
run;
```

```
data effect;  
  set bridge;  
  l_hi=log10(hi);  
run;
```

```
proc sort data=effect;  
  by pid day run;  
run;
```

```
/******Estimation des GMT non corrigés*****/
```

```
proc corr data=nobridge cov;  
  var run group_nb day;  
run;  
quit;
```

```
*On montre ici que day et run sont fortement corrélés;  
*donc l'effet day est mélangé avec l'effet run ;
```

```

proc mixed data=nobridge;
  class pid group_nb day;
  model l_hi=group_nb day group_nb*day/outp=residus residuals solution covb;
  repeated day/subject=pid r type=un;
  lsmeans group_nb*day/cl;
  ods output lsmeans=gmt_non_corrige;
run;
quit;

proc sql;
  create table res as
  select resid
  from residus;
quit;

proc univariate data=res normaltest;
  qqplot resid/normal;
run;

data gmt_non_corrige;
  set gmt_non_corrige;
  gmt=10**estimate;
  ll=10**lower;
  ul=10**upper;
run;

ODS HTML FILE='gmt non corrigé.xls';
  proc print data=gmt_non_corrige;
    var group_nb day gmt ll ul;
  run;
ODS HTML CLOSE;

/*****Estimation des variances de delta run et des gmt*****/

  /*Variance delta run*/

  proc transpose data=effect out=delta_r;
  var l_hi;
  by pid;
  run;

  data delta_r;
  set delta_r;
  d_run=COL1-COL2;
  rename COL1=l_hi_r1;
  rename COL2=l_hi_r2;
  run;

  proc means data=delta_r var vardef=df;
  var d_run;
  output out=variance_d_run var(d_run)=variance;
  run;

  proc univariate data=delta_r;
  var d_run;
  run;

```

```

/*Variance gmt non corrigés*/

proc mixed data=nobridge;
  class pid group_nb day;
  model l_hi=group_nb day group_nb*day/solution;
  repeated day/r subject=pid type=un;
  lsmeans group_nb*day/cov;
  ods output lsmeans=gmtb;
run;
quit;

/*****Estimation de l'effet run et correction*****/

/*Estimation de l'effet run*/

proc corr data=effect cov;          *On montre ici que group_nb et day sont independants
  var run group_nb day;           de run dans le bridge ;
run;

proc mixed data=effect;
  class group_nb day run pid;
  model l_hi=day run/solution outp=pred;
  repeated run/r rcorr subject=pid type=cs;
  lsmeans run/cov diff;
  ods output solutionf=solf;
run;
quit;

/*Vérification des hypothèses sur les résidus*/

  *Additivité du modèle*;

goptions reset=all;
goptions htext=1.5;

  symbol1 v=plus line=1 i=join;
  symbol2 v=X line=2 i=join;
  symbol3 v=star line=41 i=join;
  axis1 label=(angle=90 'log_10(HI)');
  axis2 label=('Run') order=(0.5 to 2.5 by 0.5);

proc gplot data=effect;
  title'Nuage de points de log_10(hi) et run';
  plot l_hi*run=pid /vaxis=axis1 haxis=axis2;
run;
quit;

  *Normalité des résidus*;

proc sql;
  create table res as
  select resid
  from pred;

```

```

quit;

proc univariate data=res normaltest;
  qqplot resid/normal;
run;

  *Homoscédasticité des résidus*;

proc glm data=pred;
  class day pid run;
  model resid=run;
  means run/hovtest=bartlett;
run;
quit;

goptions reset=all;
symbol v=dot;
axis1 label=(angle=90 'résiduals');
axis2 label=('Run') order=(0.5 to 2.5 by 0.5);

proc gplot data=pred;
  title'Nuage de points de residuals et run';
  plot resid*run=pid/vref=0 vaxis=axis1 haxis=axis2;
run;
quit;

/*Correction*/

proc sql;
  create table two as
  select one.*, d_run
  from one, (select distinct estimate as d_run from solf where
effect='run' and run=1);
quit;

data three;
set two;
if run=1 then
  l_hi=l_hi-d_run/2;
  else
  l_hi=l_hi+d_run/2;
run;

  /******Estimation des gmt après correction*****/

proc sort data=three;
  by pid group_nb day;
run;

proc univariate data=three noprint;
  var l_hi;
  output out=four mean=l_hi;
  by pid group_nb day;
run;

proc sort data=four;
  by pid day;

```

```

run;

proc mixed data=four;
  class pid group_nb day;
  model l_hi=group_nb day group_nb*day/solution;
  repeated day/ subject=pid type=un;
  lsmeans group_nb*day/cl;
  ods output lsmeans=gmt_corrige;
run;
quit;

data gmt_corrige;
  set gmt_corrige;
  gmt=10**estimate;
  ll=10**lower;
  ul=10**upper;
run;

ODS HTML FILE='gmt corrigé sans rectif var.xls';
  proc print data=gmt_corrige;
    var group_nb day gmt ll ul;
  run;
ODS HTML CLOSE;

/*****Evaluation des différentes variabilités inter effet*****/

proc mixed data=one;
  class group_nb day run pid;
  model l_hi=group_nb day/s outp=pred2;
  random run day*run/v s;
  repeated day/subject=pid type=cs r;
  ods output covparms=effet;
run;

  /***pour effet pid***/
proc mixed data=one;
  class group_nb day run pid;
  model l_hi=group_nb day group_nb*day/s outp=pred3;
  repeated day/subject=pid type=cs r;
  ods output covparms=effet_pid;
run;

  /*Vérification des hypothèses sur les résidus*/

  *Normalité des résidus*;

proc sql;
  create table res2 as
  select resid
  from pred2;
quit;

proc univariate data=res2 normaltest;
  qqplot resid/normal;
run;

```



```

*Homoscédasticité des résidus*;

proc glm data=pred2;
  class day pid run;
  model resid=run;
  means run/hovtest=bartlett;
run;

goptions reset=all;
  symbol v=dot;

proc gplot data=pred2;
  plot resid*run=pid;
run;
quit;

/*****Création tableau général*****/

data gmtb;
  set gmtb;
  rename estimate=estimateold;
run;

proc sort data=gmtb;
  by group_nb day;
run;

proc sort data=gmt_corrige;
  by group_nb day;
run;

data gmtf;
  merge gmt_corrige gmtb;
  by group_nb day;
run;

proc sql;
  create table variance as
  select gmtf.*, variance_d_run
  from gmtf, (select variance as variance_d_run from variance_d_run);
run;

data variance;
  set variance;
  variance_d_run=variance_d_run/29;
  n_row=_n_;
run;

data variance;
  set variance;
  array x(6) cov1 cov2 cov3 cov4 cov5 cov6;
  do i=1 to 6;
    variance_gmt=x(i);output;
  end;
run;

```

```

data variance;
  set variance;
  if i=n_row then output;
  else delete;
run;

data variance;
  set variance;

  interrune=0.000976;          *obtenu avec day et group_nb fixe et run et day*run aléatoire;
  varresid=0.007570;

  if day=35 then
    covariance= -interrune;
  if day=0 then do;
    if group_nb=1 then
      covariance= interrune+(3/928)*varresid;
    else
      covariance= interrune+(1/638)*varresid;
  end;
  if day=21 then do;
    if group_nb=1 then
      covariance= interrune+(6/928)*varresid;
    else
      covariance= interrune+(4/957)*varresid;
  end;
run;

data variance;
  set variance;
  rho=covariance/(sqrt(variance_gmt)*sqrt(variance_d_run));
  if day ~= 35 then
    variance_tot=variance_gmt+(variance_d_run/4)- covariance;
  else
    variance_tot=variance_gmt+(variance_d_run/4)+ covariance;

  d_f=(variance_tot)**2/(((variance_d_run/4)**2)/28)+((variance_gmt)**2/128)
  +((rho**2)*((variance_d_run/4)/28)*((variance_gmt)/128)*(28+128+1/2))+2*rho
  *sqrt(variance_gmt)*sqrt(variance_d_run/4)*(((variance_d_run/4)/28)+((variance_gmt)/128)); /*Equation de Welch-Satterthwaite*/
  q=quantile('T',97.5/100,d_f);
  ll_gmt=estimate-q*sqrt(variance_tot);
  ul_gmt=estimate+q*sqrt(variance_tot);
run;

data variance;
  set variance;
  gmt=10**estimate;
  ll=10**ll_gmt;
  ul=10**ul_gmt;
run;

ODS HTML FILE='Tableau recap variances et gmt.xls';
proc print data=variance;
  var effect group_nb day gmt ll ul covariance variance_tot rho d_f q;
run;

```

```
ODS HTML CLOSE;
```

```
/******Graphiques******/
```

```
/******Estimation des paramètres de la fonction*****/
```

```
proc nlin data=variance;  
  parms b1=0.9 b2=2 b4=0.09;  
  bounds 0<b4<0.5;  
  bounds b2>0;  
  if (day<=21) then  
    model estimate=b1*(1+b2)/(1+b2*exp(-(b2/5)*day));  
  else model estimate=(b1*(1+b2)/(1+b2*exp(-(b2/5)*21)))*(1+b4)/(1+b4*exp(-b4*(day-21)));  
  by group_nb;  
  ods output ConvergenceStatus=cs;  
  ods output ParameterEstimates=est;  
run;
```

```
proc nlin data=variance;  
  parms b1=0.8 b2=2 b4=0.09;  
  bounds 0<b4<0.5;  
  bounds b2>0;  
  if (day<=21) then  
    model l1_gmt=b1*(1+b2)/(1+b2*exp(-(b2/5)*day));  
  else model l1_gmt=(b1*(1+b2)/(1+b2*exp(-(b2/5)*21)))*(1+b4)/(1+b4*exp(-b4*(day-21)));  
  by group_nb;  
  ods output ConvergenceStatus=cs2;  
  ods output ParameterEstimates=est2;  
run;
```

```
proc nlin data=variance;  
  parms b1=1 b2=2 b4=0.09;  
  bounds 0<b4<0.5;  
  bounds b2>0;  
  if (day<=21) then  
    model ul_gmt=b1*(1+b2)/(1+b2*exp(-(b2/5)*day));  
  else model ul_gmt=(b1*(1+b2)/(1+b2*exp(-(b2/5)*21)))*(1+b4)/(1+b4*exp(-b4*(day-21)));  
  by group_nb;  
  ods output ConvergenceStatus=cs3;  
  ods output ParameterEstimates=est3;  
run;
```

```
ODS HTML FILE='variance corrigé.xls';
```

```
proc print data=variance;  
  var group_nb day gmt l1 ul;
```

```
run;
```

```
ODS HTML CLOSE;
```

```
/******Approximation des fonctions*****/
```

```
proc sql noprint;  
  select estimate into: b11 from est where group_nb=1 and parameter='b1';  
  select estimate into: b21 from est where group_nb=1 and parameter='b2';
```

```

select estimate into: b41 from est where group_nb=1 and parameter='b4';
select estimate into: b12 from est where group_nb=2 and parameter='b1';
select estimate into: b22 from est where group_nb=2 and parameter='b2';
select estimate into: b42 from est where group_nb=2 and parameter='b4';
select estimate into: b13 from est2 where group_nb=1 and parameter='b1';
select estimate into: b23 from est2 where group_nb=1 and parameter='b2';
select estimate into: b43 from est2 where group_nb=1 and parameter='b4';
select estimate into: b14 from est2 where group_nb=2 and parameter='b1';
select estimate into: b24 from est2 where group_nb=2 and parameter='b2';
select estimate into: b44 from est2 where group_nb=2 and parameter='b4';
select estimate into: b15 from est3 where group_nb=1 and parameter='b1';
select estimate into: b25 from est3 where group_nb=1 and parameter='b2';
select estimate into: b45 from est3 where group_nb=1 and parameter='b4';
select estimate into: b16 from est3 where group_nb=2 and parameter='b1';
select estimate into: b26 from est3 where group_nb=2 and parameter='b2';
select estimate into: b46 from est3 where group_nb=2 and parameter='b4';
quit;

```

```

data tmp;
  set variance;
  l_hi1=NULL;
  l_hi2=NULL;
  l_hi3=NULL;
  l_hi4=NULL;
  l_hi5=NULL;
  l_hi6=NULL;
  hi1=NULL;
  hi2=NULL;
  hi3=NULL;
  hi4=NULL;
  hi5=NULL;
  hi6=NULL;

```

run;

```

data tmp2;
  set variance;

  do day=0 to 35;
    group_nb=3;

    if (day<=21) then
      l_hi1=&b11*(1+&b21)/(1+&b21*exp(-(&b21/5)*day));
    else l_hi1=(&b11*(1+&b21)/(1+&b21*exp(-
(&b21/5)*21)))*(1+&b41)/(1+&b41*exp(-(&b41)*(day-21)));

    hi1=10**l_hi1;
    gmt=null;
    output;
  end;

  do day=0 to 35;
    group_nb=4;

    if (day<=21) then
      l_hi2=&b12*(1+&b22)/(1+&b22*exp(-(&b22/5)*day));
    else l_hi2=(&b12*(1+&b22)/(1+&b22*exp(-
(&b22/5)*21)))*(1+&b42)/(1+&b42*exp(-(&b42)*(day-21)));

    hi2=10**l_hi2;
    hi1=null;

```

```

        output;
    end;

do day=0 to 35;
    group_nb=5;

        if (day<=21) then
            l_hi3=&b13*(1+&b23)/(1+&b23*exp(-(&b23/5)*day));
        else l_hi3=(&b13*(1+&b23)/(1+&b23*exp(-
(&b23/5)*21)))*(1+&b43)/(1+&b43*exp(-(&b43)*(day-21)));

        hi3=10**l_hi3;
            hi2=null;
            output;
        end;

do day=0 to 35;
    group_nb=6;

        if (day<=21) then
            l_hi4=&b14*(1+&b24)/(1+&b24*exp(-(&b24/5)*day));
        else l_hi4=(&b14*(1+&b24)/(1+&b24*exp(-
(&b24/5)*21)))*(1+&b44)/(1+&b44*exp(-(&b44)*(day-21)));

        hi4=10**l_hi4;
            hi3=null;
            output;
        end;

do day=0 to 35;
    group_nb=7;

        if (day<=21) then
            l_hi5=&b15*(1+&b25)/(1+&b25*exp(-(&b25/5)*day));
        else l_hi5=(&b15*(1+&b25)/(1+&b25*exp(-
(&b25/5)*21)))*(1+&b45)/(1+&b45*exp(-(&b45)*(day-21)));

        hi5=10**l_hi5;
            hi4=null;
            output;
        end;

do day=0 to 35;
    group_nb=8;

        if (day<=21) then
            l_hi6=&b16*(1+&b26)/(1+&b26*exp(-(&b26/5)*day));
        else l_hi6=(&b16*(1+&b26)/(1+&b26*exp(-
(&b26/5)*21)))*(1+&b46)/(1+&b46*exp(-(&b46)*(day-21)));

        hi6=10**l_hi6;
            hi5=null;
            output;
        end;
run;

/*****Représentation Graphique*****/

proc append data=tmp2 base=tmp force;
run;

```

```

data tmp;
  set tmp;
  if group_nb=1 then
    gmt1=gmt;
  else
    gmt2=gmt;
run;

proc sort data=tmp;
  by group_nb day;
run;

goptions htext=1.5;
  symbol1 v=dot i=none color=blue;
  symbol2 v=dot i=none color=orange;
  symbol3 v=none i=join color=green;
  symbol4 v=none i=join color=purple;
  symbol5 v=none i=join color=black;
  symbol6 v=none i=join color=grey;
  symbol7 v=none i=join color=black;
  symbol8 v=none i=join color=grey;
  axis1 label=(angle=90 'GMT') logbase=10 logstyle=expand
order=(1,10,100,1000);
  axis2 label=('Day');
  axis3 label=(angle=90 'GMT') order=(0 to 700 by 100);
  legend1 value=(tick=1 'GMT groupe1' tick=2 'GMT groupe2' tick=3 'Groupe1'
tick=4 'Groupe2' tick=5 'CI inf Gr1' tick=6 'CI inf Gr2' tick=7 'CI sup
Gr1' tick=8 'CI sup Gr2')
    label=none
    frame;

proc gplot data=tmp;
  title'Evolution GMT corrigés avec correction de la variance';
  plot (gmt1 gmt2 hi1 hi2 hi3 hi4 hi5 hi6)*day / overlay vaxis=axis1
haxis=axis2 legend=legend1;
run;
quit;

proc gplot data=tmp;
  title'Evolution GMT corrigés avec correction de la variance';
  plot (gmt1 gmt2 hi1 hi2 hi3 hi4 hi5 hi6)*day / overlay vaxis=axis3
haxis=axis2 legend=legend1;
run;
quit;

```