Pharmacogénétique en Pharmacocinétique de population : tests et sélection de modèles

Julie Bertrand

UNIVERSITÉ PARIS.DIDEROT (Paris 7)

ÉCOLE DOCTORALE : 393
SANTÉ PUBLIQUE : ÉPIDÉMIOLOGIE ET SCIENCES DE L'INFORMATION BIOMÉDICALE

DOCTORAT

Julie Bertrand

Pharmacogénétique en Pharmacocinétique de population : tests et sélection de modèles

Thèse dirigée par le Pr. France Mentré

Soutenue le 1er décembre 2009

JURY

Monsieur Jean-Yves Mary	Président du Jury
Monsieur Michel Tod	Rapporteur
Madame Hélène Jacqmin-Gadda	Rapporteur
Madame France Mentré	Directeur de thèse
Madame Emmanuelle Comets	Co-encadrante
Madame Marylore Chenel	Co-encadrante
Madame Caroline Solas	Examinateur

Jamais assez de remerciements…

A Jean-Yves Mary pour avoir accepté de participer à mon jury de thèse et m'avoir accompagnée dans les méandres administratives qui entourent la soutenance d'une thèse dans l'université Paris Diderot.

A Michel Tod et Hélène Jacqmin-Gadda pour avoir accepté d'être les rapporteurs de ma thèse. Vos commentaires m'ont permis d'améliorer ce manuscrit et m'ont ouvert de nouvelles perspectives.

A Caroline Solas, pour avoir accepté d'être examinateur dans mon jury et apporter ainsi son regard expert sur les aspects cliniques et/ou méthodologiques propre à la pharmacogénétique.

A ma directrice de thèse, France Mentré pour votre accueil tout d'abord puis votre curiosité et vos connaissances ainsi que votre aisance à les transmettre. Votre enthousiasme et votre rigueur ont su créer un cadre plus que propice à la recherche pour des étudiants, des chercheurs et des ingénieurs. De nombreuses personnes riches d'expériences et d'idées avec lesquelles j'ai pu interagir tout au long de ma thèse.

A mon encadrante, Emmanuelle Comets pour nos échanges scientifiques et le soutien moral constant durant ces trois années. Pour ce qui est de l'habileté à diriger des recherches, il ne te manque que le titre en ce qui me concerne ;-)

A Céline Lafont qui m'a chaleureusement accueillie et encadrée dans le département de pharmacocinétique clinique de l'institut de recherches internationales Servier, institut ayant financé ma thèse et à Marylore Chenel qui a pris la relève avec dynamisme. Vous avez toutes les deux éclairé de votre expérience tant mes travaux appliqués que méthodologiques et vous m'avez permis de découvrir et apprécier la recherche dans l'industrie pharmaceutique.

A tous les membres de l'UMR 738 INSERM Paris Diderot 738 sur le site de la faculté, dans les locaux de l'hôpital et même un peu plus loin.

A Caroline, Corine, Pierre, Rada et plus récemment Steven et Thu qui m'ont supporté au quotidien parce que nous partagions le même bureau. A Anne, George, Hervé, Houda, Jérôme, Quentin, Sylvie et Thu-thuy qui étaient à peine mieux protégés de mes sautes d'humeur parce que dans des bureaux adjacents. Vous avez tous participé à faire de ces trois années une magnifique expérience humaine.

A tous mes amis du sud de la France jusqu'à Paris, pour leur compréhension et leur soutien, avec une mention spéciale pour mes relecteurs Aude et Loïc.

A ma famille, pour cette force qui me porte et ne vacille jamais. Je ne vous serai jamais assez reconnaissante...

A Cyril qui est tout simplement parfait et qui brille !

Table des matières

Productions scientifiques liées à la thèse :

Articles publiés

Bertrand J., Comets E., Laffont C.M., Chenel M., Mentré F., Pharmacogenetics and population pharmacokinetics : impact of the design on three tests using the SAEM algorithm. *Journal of Pharmacokinetics and Pharmacodynamics*, 2009. vol. 36, p. 317-339

Bertrand J., Panhard X., Tran A., Rey E., Auleley S., Duval X., Salmon D., Tréluyer J.M., Mentré F., the COPHAR2- ANRS 111 study group. Influence of pharmacogenetics on indinavir disposition and short-term response in HIV patients initiating HAART. *European Journal of Clinical Pharmacology*, 2009, vol. 65, p. 667-78

Bertrand, J., Comets, E., Mentré, F. Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *Journal of Biopharmaceutical Statistics*, 2008, vol. 18, p. 1084-1102

Articles en préparation

Bertrand J., Laffont C.M., Comets E., Chenel M., Mentré F. Test for genetic covariate in a complex parent-metabolite model. En préparation pour *The American Association of Pharmaceutical Scientists Journal*

Bertrand J., Comets E., Chenel M., Mentré F. Alternatives to asymptotic tests for the analysis of pharmacogenetic data using nonlinear mixed effects models. En préparation pour *Biometrics*

Communications orales

Bertrand J., Comets E., Laffont C.M., Chenel M., Mentré F. Model-based tests to detect gene effect in pharmacokinetic studies. *18^{th} Population Group Approach in Europe conference*, St-Pertersbourg, Russie, Juin 2009.

Bertrand, J., Comets E., Laffont C.M., Chenel M., Mentré F. Place du test de permutation dans les modèles non linéaires à effets mixtes : application en pharmacogénétique. *$41^{èmes}$ Journées de Statistique*, Bordeaux, France, Mai 2009.

Posters

Chou M., Bertrand J., Segeral O., Verstuyft C., Borand L., Comets E., Becquemont L., Ouk V., Mentré F., Taburet A.M., the ANRS-12154 study group. Inter- and intrapatient variabilities in nevirapine plasma concentrations in HIV-infected Cambodian patients and the effect of CYP2B6 genetic polymorphism : ANRS

12154 Study. *16th Conference on Retroviruses and Opportunistic Infections*, Montréal, Canada, Février 2009.

Bertrand J., Comets E., Mentré F. Properties of different tests to detect the effect of a genetic covariate on pharmacokinetic parameters using the SAEM algorithm for several designs. *17th Population Group Approach in Europe conference*, Marseille, France, Juin 2008.

Bertrand J., Comets E., Mentré F. Model-based tests of the effect of a genetic covariate on pharmacokinetic parameters using the SAEM algorithm. *American Conference of Pharmacometrics*, Tucson, États-Unis, Mars 2008.

Bertrand J., Panhard X., Tran A., Rey E., Auleley S., Duval X., Salmon D., Tréluyer J.M., Mentré F., the COPHAR 2-ANRS 111 study group. Influence of pharmacogenetic on pharmacokinetic interindividual variability of indinavir and lopinavir in HIV patients (COPHAR 2-ANRS 111 trial). *16th Population Group Approach in Europe conference*, Copenhague, Danemark, Juin 2007.

Résumé publié

Bertrand J. Analyse de l'information génétique en pharmacocinétique de population. *Revue d'Épidémiologie et de Santé Publique*, 2008, vol. 56, p. 220-221

Chapitre 1

Introduction

Ce travail de thèse porte sur les méthodes de tests et les stratégies de sélection de modèles utilisées dans le cadre de l'analyse de données pharmacogénétiques en pharmacocinétique de population. De plus en plus d'études sont réalisées dans ce domaine et différents choix de protocoles ont été adoptés. Une première classe d'études reposait sur le concept "d'échantillon enrichi" où les sujets étaient sélectionnés par leur génotype afin d'assurer une représentation homogène (Comets et al., 2007). Intrinsèquement, ces études ne permettaient pas d'explorer un nombre important de polymorphismes. Afin d'augmenter le nombre de polymorphismes pouvant être analysés, les études pharmacogénétiques sont désormais intégrées dans des essais plus larges où les génotypages peuvent être par exemple réalisés en routine clinique. Dans ces nouvelles études, la distribution des polymorphismes n'est plus contrôlée, les populations et leur traitement sont plus hétérogènes. Les données pharmacocinétiques sont généralement moins riches et les méthodes d'analyse classiques ne sont plus adaptées. Il est nécessaire d'avoir recours à des outils statistiques plus puissants comme les modèles non linéaires à effets mixtes (MNLEM).

Dans la partie 1.1, nous définissons la notion de pharmacogénétique ainsi que son rôle dans le développement et l'utilisation thérapeutique du médicament. Dans les parties 1.2 et 1.3, nous décrivons brièvement ses composantes pharmacocinétiques et génétiques. Puis dans les parties 1.4 et 1.5, nous présentons les méthodes classiques et l'approche dite "de population" pour l'analyse des données pharmacogénétiques. Enfin dans la partie 1.6, nous présentons les objectifs de la thèse dans ce contexte.

1.1 La pharmacogénétique

1.1.1 Définition

La pharmacogénétique est une composante de la pharmacogénomique ; alors que la pharmacogénomique étudie comment les variations de l'ensemble du génome peuvent expliquer la variabilité de la réponse à un médicament entre plusieurs individus,

la pharmacogénétique étudie les variations des gènes codants pour des protéines impliquées dans l'absorption, la distribution, le métabolisme, l'élimination et l'effet du médicament (Licinio et Wong, 2002; Kalow et al., 2001).

Les autorités de santé en Europe et aux États-Unis ont statué récemment sur la place des études pharmacogénétiques au cours du développement de médicaments (for human medicinal products, 2008; FDA, 2008). Ce domaine s'est développé il y a désormais plus de 50 ans (Motulsky, 1957) et depuis 10 ans le nombre d'études pharmacogénétiques n'a cessé d'augmenté comme l'illustre la figure 1.1 où l'évolution du nombre d'études pharmacogénétiques est représentée de début 2000 à septembre 2009 par un graphique en barres. Cette sélection a été réalisée par mots clés dans la base de donnée MEDLINE (PubMed) sur le titre et le résumé, en se limitant aux études réalisées chez l'homme et publiées en langue anglaise pendant cette période. Ces études ne portent pas toutes sur la détection d'un polymorphisme génétique mais aussi sur l'impact économique du génotypage ou encore sur des interactions médicamenteuses dans une population génétiquement sélectionnée.

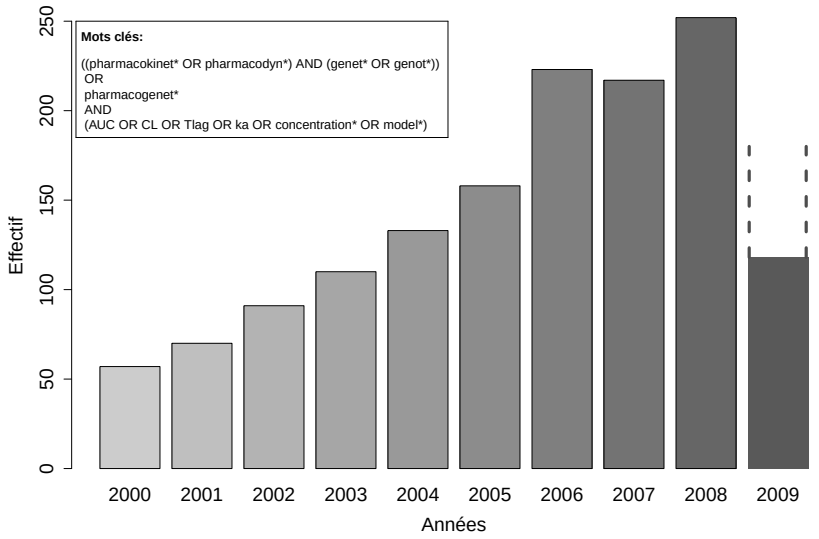


Fig. 1.1 – Évolution du nombre d'études pharmacogénétiques répertoriées dans la base de données MEDLINE (PubMed) au 21 septembre 2009

1.1.2 Rôle dans le développement et l'utilisation thérapeutique du médicament

Un certain nombre d'étapes jalonnent le développement d'un médicament. Dans les essais de phase I réalisés habituellement chez le volontaire sain, les études pharmacocinétiques et pharmacodynamiques permettent de déterminer la dose maximale

tolérée. Dans les essais de phase II réalisées chez le patient avec un nombre plus important de sujets, les études pharmacocinétiques et pharmacodynamiques permettent d'évaluer l'efficacité et l'influence de covariables sur la relation dose-réponse. Les études pharmacogénétiques dans ces premières phases du développement permettent d'évaluer, pour des protéines clés, l'impact de polymorphismes génétiques sur la réponse au médicament et le cas échéant d'influer sur la sélection des patients à l'étape suivante : les essais de phase III. Ces derniers sont des étapes clés du développement du médicament qui impliquent de très larges effectifs. A ce stade, les études pharmacogénétiques permettent de considérer un plus grand nombre de polymorphismes et de quantifier la part de la variabilité inter-sujets associée à ces polymorphismes.

Dix à quinze ans peuvent s'écouler entre la découverte d'une nouvelle molécule et sa mise sur le marché. Si cette durée reste relativement stable au cours du temps, le nombre de nouvelles molécules mises sur le marché décroît (Kuhlmann, 1999) et certaines molécules efficaces sont retirées pour cause d'effet secondaires. La recherche en pharmacogénétique peut potentiellement augmenter les chances pour les nouvelles molécules d'être validées par les autorités de santé, par exemple en éliminant les molécules interagissant avec des protéines de l'organisme fortement polymorphes. Des études pharmacogénétiques peuvent aussi permettre de définir génétiquement une population plus susceptible de bénéficier de certains traitements (Tanigawara et al., 1999).

Après la mise sur le marché, les études pharmacogénétiques permettent d'améliorer la compréhension de la relation entre l'efficacité et/ou la toxicité du médicament et les polymorphismes. L'impact des polymorphismes est aussi étudié sur des interactions médicamenteuses non évaluées lors du développement du médicament (Chen et al., 2009) ou encore sur la survenue d'effets indésirables graves rares ou tardifs qui ne peuvent être observés dans la durée des études classiques.

Dès 1920, des variations inter-ethniques ont été observés dans la réponse aux médicaments (Paskind, 1921). Ces variations s'expliquent en partie par la diversité génétique entre les populations (Wilson et al., 2001). Des études pharmacogénétiques peuvent donc participer à l'exportation des traitements développés en Europe vers d'autres populations, notamment en guidant le choix de la dose à administrer aux patients.

1.2 La pharmacocinétique

1.2.1 Un processus physiologique complexe

La pharmacocinétique étudie le devenir du médicament dans l'organisme (Gabrielsson et Weiner, 1999). Ce processus peut être résumé en quatre grandes étapes : Absorption, Distribution, Métabolisation et Élimination (ADME). La figure 1.2 illustre dans le corps humain les différents sites d'administration (intra-artérielle, intra-veineuse, intra-musculaire/sous-cutanée, pulmonaire et orale), la distribution à travers le

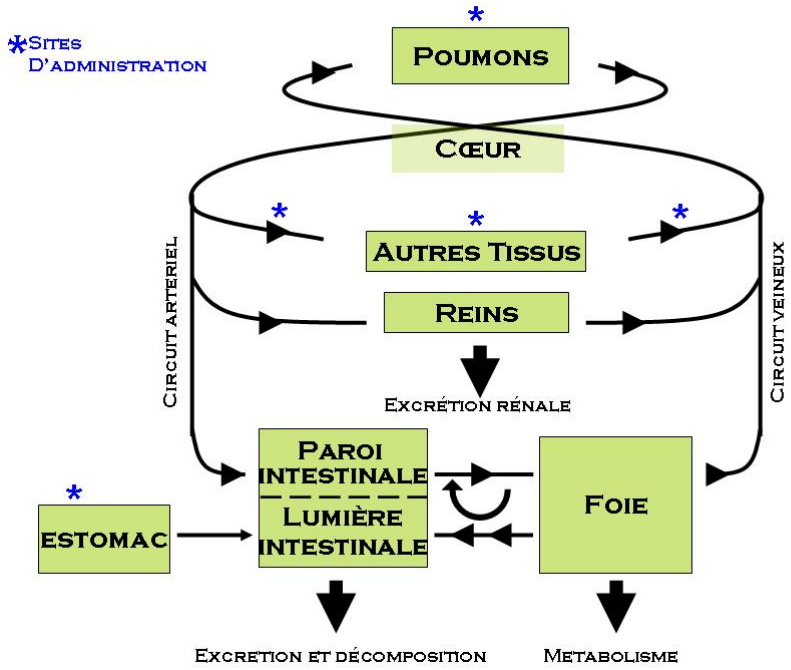


FIG. 1.2 – Parcours d'un médicament dans le corps humain
Source : premier cours *Review of basic principles* du séminaire *The Sheiner/Rowland advanced course in pharmacokinetics/pharmacodynamics*

circuit sanguin, le cycle entéro-hépatique et les organes d'élimination.

La phase d'absorption regroupe tous les évènements du trajet du médicament du site d'administration jusqu'au site de mesure. Sa vitesse est limitée par le flux sanguin et la perméabilité des membranes. La fraction de médicament qui rejoint le circuit plasmatique ou biodisponibilité F, est limitée par le taux de dissolution et l'effet de premier passage du aux enzymes du métabolisme rencontrées à travers l'intestin, la membrane intestinale et le foie. La distribution dans les différents tissus de l'organisme est quant à elle conditionnée par le degré de perfusion sanguine du tissu et la perméabilité des membranes tissulaires. Toutes les protéines circulant dans le sang ou fixées sur les membranes sont susceptibles de se lier au médicament et donc d'influencer sa distribution. Le processus d'élimination comprend le métabolisme à travers les enzymes du foie et l'excrétion rénale, c'est à dire la filtration glomérulaire et la sécrétion capillaire moins la réabsorption tubulaire. Enfin, les caractéristiques physicochimiques du médicament (lipophile/hydrophile, masse moléculaire, polarité et nature acide/base) déterminent son comportement à chacune de ces étapes. Les concentrations plasmatiques du médicament, après une absorption orale par exemple, augmentent durant la phase d'absorption jusqu'à atteindre un pic puis diminuent progressivement du fait de son élimination (figure 1.3).

L'aire sous la courbe des concentrations en fonction du temps (AUC) est une mesure de l'exposition au médicament comme la concentration résiduelle qui est la concentration de médicament la plus basse observée avant la prochaine prise. Cette dernière est une

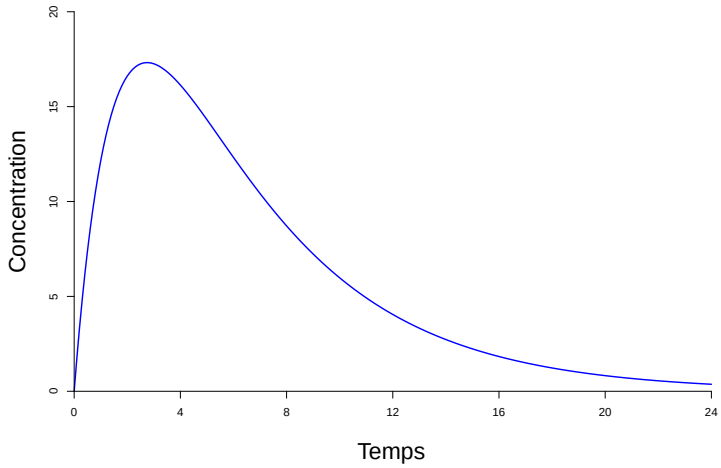


FIG. 1.3 – Exemple de profil de concentrations plasmatiques au cours du temps, après une absorption orale

mesure plus pertinente dans certaines classes thérapeutiques comme les antibiotiques. De manière générale, la marge thérapeutique d'un médicament est définie par la concentration maximale non toxique et la concentration minimale efficace. Il est d'autant plus nécessaire d'étudier les propriétés pharmacocinétiques d'un médicament que cette marge est étroite.

1.2.2 La modélisation

Elle permet d'intégrer des connaissances *a priori* sur le médicament qui ont été accumulées aux précédentes étapes du développement du médicament comme le fait que son élimination soit limitée par une enzyme ou qu'il soit susceptible de subir un fort effet de premier passage ou un mécanisme de rétro-conversion (= formation du produit parent à partir du métabolite).

Ces informations sont intégrées au modèle structural qui sert à décrire l'évolution des concentrations dans le plasma au cours du temps. Ce modèle est constitué d'un ensemble de compartiments connectés entre eux autour d'un compartiment central (= système mamillaire) ou à la suite les uns des autres (= compartiments de transit), comme illustré par la figure 1.4.

Les variations de quantité au cours du temps dans ces différents compartiments peuvent être transcrites sous forme de système d'équations différentielles, avec une équation par compartiment. Les flèches entre les compartiments correspondent aux vitesses d'absorption (du compartiment extérieur vers le compartiment de mesure) de passage (d'un compartiment à un autre) et d'élimination (du compartiment de mesure vers l'extérieur). Ces équations écrites en termes de masses peuvent être traduites en concentrations en les rapportant au volume propre à chaque compartiment. Si elle existe,

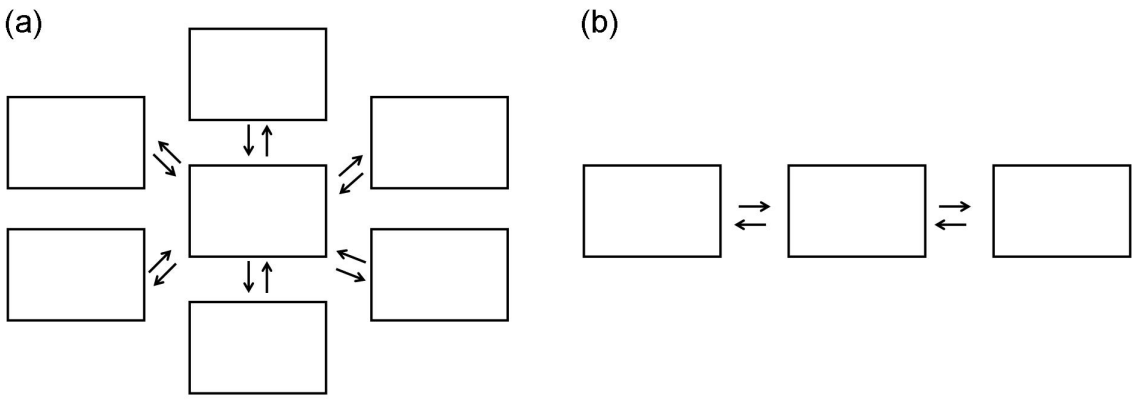


FIG. 1.4 – Schémas représentant un système mamillaire de compartiments périphériques tous reliés à un compartiment central (a) et un système de compartiments de transit ou les compartiments sont alignés à la suite les uns des autres (b).

une solution analytique de ce système peut alors être utilisée pour formuler le modèle en une fonction mathématique du temps (GIBALDI et PERRIER, 1982). Pour la librairie du logiciel MONOLIX (LAVIELLE, 2008), nous avons ainsi formulé soixante-dix-neuf modèles pharmacocinétiques de un à trois compartiments en considérant quatre modes d'administration possibles (avec ou sans délai) après une dose unique, des doses multiples ou à l'équilibre (BERTRAND et MENTRÉ, 2008).

Pour que tous les paramètres d'un modèle soient identifiables, des mesures doivent être effectuées dans chaque compartiment (plasma et urine, intra et extra-cellulaire), ce qui est rarement possible. Les paramètres du modèle sont donc estimés à une constante près (qui ne peut être estimée), cette constante peut aussi être fixée à une valeur arbitraire. Par exemple pour une administration orale, des mesures simultanées de concentration plasmatique et de la quantité sécrétée dans les urines sont nécessaires pour estimer la fraction de biodisponibilité (GABRIELSSON et WEINER, 1999). Dans cette thèse, nous avons ainsi estimé des volumes et clairances (= quantité de médicament éliminée par unité de temps, caractérisant le processus d'élimination) apparents, c'est à dire rapporté à la fraction de biodisponibilité. Pour estimer ce paramètre, il est aussi possible d'utiliser des concentrations mesurées après des doses orale et intra-veineuse administrées aux mêmes sujets.

La modélisation permet de décrire des processus physiologiques et d'analyser conjointement les observations obtenues dans les différents compartiments du modèle, par exemple les concentrations de la molécule parent et d'un métabolite actif ou encore les concentrations extra-cellulaires et intra-cellulaires.

1.2.3 La variabilité pharmacocinétique

Certains médicaments ont une variabilité pharmacocinétique très importante d'un sujet à l'autre et/ou d'une occasion à l'autre. Pour une même dose, cette variabilité peut entraîner l'absence complète d'effet thérapeutique comme la survenue d'effets indésirables graves.

Cette variabilité est documentée et une recherche des covariables susceptibles de l'expliquer est réalisée au cours du développement du médicament puis durant son utilisation thérapeutique. Les covariables classiquement considérées sont d'origine physiopathologique comme le sexe, l'âge ou le poids, d'origine environnementale comme la consommation de tabac ou d'alcool, d'origine clinique comme l'observance et enfin d'origine génétique.

1.3 Les polymorphismes génétiques

Deux individus pris au hasard dans la population humaine différent génétiquement d'environ 0.5%, 90% de ces différences sont des variations du nombre de copies de l'acide désoxyribonucléique (ADN) et les 10% restants résultent de *single nucleotide polymorphisms* (ou SNP) qui désignent les variations d'une seule paire de base du génome à un locus donné (Levy et al., 2007). Il existe aussi d'autres polymorphismes génétiques : les micro- et minisatellites ou variations dans la répétition d'un motif de dinucléotides ou de trinucléotides et les variations dans la lecture de l'ADN dont l'étude est appelée épigénétique.

La diversité génétique humaine se décompose en moyenne à 85% au sein d'une population locale, pour 7% entre les populations d'un même continent et pour 8% entre les populations de continents différents (Hinds et al., 2005). Cependant, certaines populations sont plus polymorphes que d'autres, notamment les populations africaines qui peuvent contenir jusqu'à 100% de la diversité génétique humaine (Long et Kittles, 2003).

1.3.1 Les *single nucleotide polymorphisms* (SNP)

Ce sont les marqueurs les plus utilisés en pharmacogénétique. La figure 1.5 représente trois SNPs, où deux variants alléliques co-existent. Un SNP peut être localisé dans une région non-codante comme dans une région codante du gène ou encore dans une zone intergénique (International HapMap Consortium, 2003). Les SNPs qui se retrouvent dans les régions codantes n'ont pas forcément de conséquences sur la production de la protéine du fait de la redondance du code génétique, ils sont dit "synonymes". Dans le cas contraire, ils peuvent entraîner une dénaturation de la protéine ou empêcher sa production. Ceux qui se retrouvent dans des régions non-codantes auront plutôt des conséquences sur

l'épissage, les facteurs de transcription, ou sur les séquences d'acide ribonucléique (ARN) non-codant (Strachan et Read, 1999).

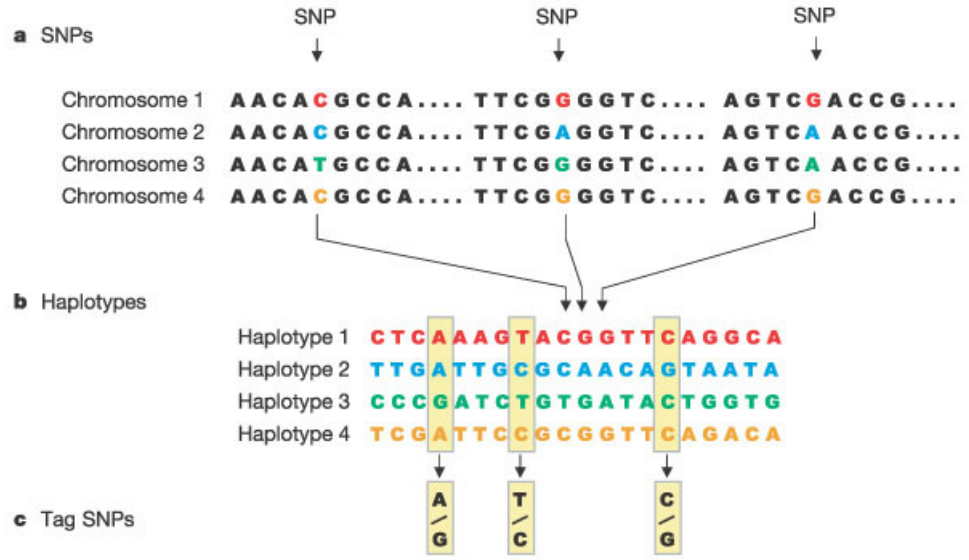


Fig. 1.5 – a) SNP : exemple d'un tronçon d'ADN observé chez 4 sujets différents. La séquence nucléotidique est globalement identique à l'exception de 3 bases où l'on observe une variation. b) Haplotypes : un haplotype est formé de la combinaison de plusieurs SNPs contigus. Ici dans une population représentative, seuls 4 haplotypes sont observés parmi l'ensemble des combinaisons possibles de 20 SNPs répartis sur 6000 bases de l'ADN dont les 3 SNPs présentés en a. c) Tag SNPs : génotyper seulement les trois SNPs encadrés permet d'identifier les 4 haplotypes, c'est à dire qu'un chromosome avec la combinaison A-T-C à ces trois SNPs possèdera de fait la combinaison des 20 SNPs correspondant à l'haplotype 1.
Source : *Science* ©AAAS

Dans le cas où il existe K formes alléliques différentes dans la population, comme l'homme est un organisme diploïde, K(K+1)/2 génotypes sont possibles : K homozygotes (= qui possède 2 copies du même allèle) et (K-1)/2 hétérozygotes (= qui possède 2 allèles différents). Le phénotype résulte de l'interaction des 2 allèles, d'une manière additive (les hétérozygotes ayant un phénotype à mi-chemin des homozygotes) ou multiplicative. Dans le cas le plus simple, un allèle (dit allèle dominant) peut complètement masquer l'effet des autres (dit allèles récessifs). La complexité du code génétique qui lui confère sa capacité d'adaptation, sa richesse et sa robustesse rend ainsi difficile toute détermination automatique du lien génotype-phénotype.

Par ailleurs, la génétique des populations enseigne que dans une large population de taille N où les accouplements sont réalisés de manière aléatoire (= les allèles paternel et maternel sont tirés au sort parmi $2 \times N$ allèles), si $p_k = p_1, ..., p_K$ sont les fréquences des k=1,...,K allèles, alors les proportions des K(K+1)/2 génotypes seront données, en une génération, par le développement de l'identité remarquable $N\left(p_1 + ... + p_K\right)^2 = Np_1^2 + 2Np_1p_2 + ... + 2Np_{K-1}p_K + Np_K^2$ dites "proportions d'Hardy-Weinberg" (Crow, 1999).

Les polymorphismes génétiques sont donc par nature des covariables multivariées dont les classes sont représentées de manière inégale dans la population. De plus, ce déséquilibre est renforcé dans le cas de polymorphismes rares, peut-être plus susceptibles d'être reliés à des variations extrêmes de la réponse (Barton, 1990).

1.3.2 Le déséquilibre de liaison

Le nombre de combinaisons de SNPs observé dans la population humaine est bien moindre que le nombre de combinaisons attendues (Daly et al., 2001). Cette réduction est due à la structuration de l'ADN en blocs haplotypiques formés par des SNPs contigus très fortement corrélés (Gabriel et al., 2002), ce qui est illustré sur la figure 1.5 b) où seulement 4 haplotypes sont observés sur les 2^{20} combinaisons possibles. De nouveau, la localisation et la longueur de ces blocs peuvent varier entre les populations.

Le déséquilibre de liaison signifie que certaines combinaisons de variants sont plus probables que d'autres. Soient 2 loci bialléliques A/a (de fréquences p/1-p) et B/b (q/1-q) où A et B sont liés préférentiellement, alors lorsque le déséquilibre est complet, pour $p \neq q$ nous aurons seulement 3 haplotypes possibles : AB, Ab et ab quand $p>q$, et si $p=q$ seulement 2 : AB et ab (alors appelé association complète). Ainsi, des effets associés à un marqueur peuvent en fait être dus à un polymorphisme non-génotypé avec lequel ce marqueur est en déséquilibre. L'information génétique étant portée par tout l'ADN, l'approche consistant à considérer un à un les différents SNPs peut sembler réductrice. C'est pourtant une étape essentielle avant de considérer les modèles multivariés de Tag SNPs (figure 1.5 c) et/ou les analyses haplotypiques pour tester des effets épistasiques ou encore des SNPs non-génotypés.

1.3.3 Principaux SNPs dans les études pharmacocinétiques

Dans les études pharmacocinétiques, les marqueurs de la variabilité génétique ont été déterminés principalement par une approche dite "gènes candidats" et portent essentiellement sur des gènes codant pour des protéines transmembranaires comme la P-glycoprotéine (P-gP) ou des enzymes du métabolisme comme le cytochrome P450.

La P-gP est une molécule de transport impliquée dans le flux de peptides endogènes et exogènes. Elle participe ainsi à l'absorption, la distribution et l'excrétion d'un nombre important de médicaments (Marzolini et al., 2003). Le gène ABCB1 codant pour la P-GP a été découvert du fait de sa sur-expression dans des souches de cellules tumorales résistantes à de nombreux agents cytotoxiques utilisés en chimiothérapie. Plus de 28 SNPs ont été caractérisés sur ce gène (Sakaeda et al., 2002) et l'impact de ces polymorphismes à été démontré, entre autres, sur la réponse aux anti-rétroviraux par Fellay et al. (2002) et Solas et al. (2007).

Les enzymes hépatiques du cytochrome P450 sont impliquées dans le métabolisme dit

de phase I (métabolisation ou bio-activation des molécules médicamenteuses) de plus de 75% des médicaments (Evans et al., 1983) et 57 gènes codants pour ces diverses enzymes ont été identifiés dans le cadre du Human Genome Project (Collins et al., 2003). Les enzymes de cette famille sont très polymorphes et reliées à de très nombreux cas d'effets indésirables (Nelson et al., 1996).

Pour l'enzyme CYP 2D6 qui est impliquée dans le métabolisme de 25% des médicaments utilisés en clinique, plus de 90 variants allèliques ont déjà été caractérisés (Daly et al., 1996). Il existe 4 catégories de phénotypes pour cette enzyme : les métaboliseurs ultra-rapides, rapides, normaux et lents. L'assignation à une de ces catégories, auparavant réalisée à l'aide de sondes métaboliques comme la debrisoquine ou la sparteine (Evans et al., 1983) est désormais reliée à la possession d'un ensemble de variants du gène *CYP*2 D6 (Owen et al., 2009).

L'enzyme CYP 3A4 est le membre du cytochrome le plus présent dans le foie avec plus de 30 SNPs identifiés à ce jour sur son gène. L'étude de Ozdemir et al. (2000) a démontré que la variabilité génétique du gène *CYP* 3A4 est susceptible d'expliquer en partie la variabilité observée sur la réponse aux substrats de cette enzyme. Mais les polymorphismes du gène *CYP* 3A4 pourraient n'être, en fait, que des marqueurs de la variabilité des polymorphismes fonctionnels du gène *CYP* 3A5 en conséquence de l'important déséquilibre de liaison entre ces deux gènes (Lamba et al., 2002).

1.3.4 La diversité génétique entre les populations

Les fréquences alléliques pour un polymorphisme donné différent d'une population à l'autre (Quaranta et al., 2006). Ces variations reflètent un héritage différentiel au cours de l'évolution de polymorphismes répartis également dans la population originelle (Patin et al., 2006). Par exemple la fréquence des métaboliseurs lents pour le CYP 2D6 est plus faible chez les populations asiatiques que dans les populations caucasiennes et la fréquence du variant C3435T du gène ABCB1 est de respectivement 80% et 50% dans les populations africaines et caucasiennes (Marzolini et al., 2003). Pourtant une telle classification est insuffisante, les fréquences alléliques pouvant différer de façon importante entre les populations constituant ces grandes classes (Ng et al., 2008). Cette variation des fréquences alléliques d'une population à l'autre nécessite de toujours travailler sur une population homogène. Dans une étude pharmacogénétique lorsque une association ethnie/pharmacocinétique du médicament à l'étude est documentée et en présence d'un mélange de populations, une stratification sur l'ethnie permet d'assurer de la validité de l'association trouvée entre la réponse au médicament et le polymorphisme à l'étude.

1.4 Les méthodes d'analyse classiques en pharmacogénétique

Dans la recherche bibliographique présentée en partie 1.1.1, nous nous sommes intéressés plus précisément à 79 articles publiés entre début 2008 et septembre 2009 qui décrivaient l'influence d'un polymorphisme génétique sur la pharmacocinétique d'un médicament. Dans environ 80% de ces articles, soit la concentration maximale ou résiduelle observée était comparée directement entre les génotypes soit une approche non compartimentale était utilisée.

Cette approche ne requiert pas d'hypothèses sur le modèle mais nécessite de réaliser un nombre important de prélèvements. L'AUC est calculée comme la somme des aires de chacun des trapèzes formés par deux concentrations observées à des temps consécutifs (Gabrielsson et Weiner, 1999). Lorsque l'hypothèse d'une décroissance mono-exponentielle entre les mesures est considérée, une méthode de calcul logarithmique peut être utilisée dans les phases d'élimination, ce qui diminue le risque de surestimation. L'extrapolation à l'infini s'appuie sur la constante d'élimination terminale qui est estimée par régression sur les dernières log-concentrations observées (3 à 4 points). A partir de l'AUC et de la constante d'élimination terminale sont obtenus le volume de distribution, la clairance et la demi-vie qui est le temps nécessaire pour diminuer de moitié la concentration plasmatique.

Cette analyse est effectuée chez chaque sujet de l'étude séparément. Les paramètres pharmacocinétiques individuels ainsi estimés sont comparés entre les différents génotypes pour le polymorphisme à l'étude par une analyse de variance à un facteur (ANOVA) comme dans l'étude de Werner et al. (2008) ou par un modèle de régression comme dans l'étude de Han et al. (2009).

Il est aussi possible de réaliser une régression non linéaire sur les concentrations de chaque sujet afin d'obtenir les paramètres individuels, en utilisant les modèles décrits dans la partie 1.2.2. Cette approche ne permet cependant pas de tenir compte des erreurs d'estimation des paramètres et, comme l'analyse non-compartimentale, requiert un nombre important de prélèvements par sujet.

1.5 L'approche de population en pharmacogénétique

Sur les 79 articles décrits plus tôt, moins de 20 ont analysé les profils de concentrations par des modèles non linéaires à effets mixtes (MNLEM) qui permettent de réduire le nombre de prélèvements en considérant globalement les concentrations de tous les patients.

1.5.1 Les modèles non linéaires à effets mixtes

Les concentrations $y_{i,j}$ du sujet i observées aux temps $t_{i,j} = 1, ...n_{i,j}$ sont décrites par une fonction $f(t_{i,j}; \phi_i)$ non linéaire. Le vecteur de paramètres ϕ_i se décompose en un vecteur d'effets fixes μ, une matrice de covariables A_i avec un vecteur de coefficients d'effets β et un vecteur d'effets aléatoires individuels η_i qui suit une loi gaussienne de moyenne nulle et de variance une matrice définie positive Ω. Un modèle exponentiel pour les effets aléatoires peut être utilisé qui permet d'assurer la positivité des estimateurs des paramètres pharmacocinétiques. Les vecteurs d'effets aléatoires et d'erreurs résiduelles sont identiquement et indépendamment distribués, la variance des erreurs résiduelles pouvant être constante $g(t_{i,j}; \phi_i) = a$, proportionnelle $g(t_{i,j}; \phi_i) = bf(t_{i,j}; \phi_i)$ ou encore une combinaison des deux $g(t_{i,j}; \phi_i) = a + bf(t_{i,j}; \phi_i)$. Le vecteur θ des paramètres de population est constitué des éléments du vecteur μ et de la matrice Ω et des coefficients du modèle de l'erreur résiduelle a et b. Des estimations Bayésiennes empiriques des paramètres individuels du sujet i peuvent être obtenues à partir de la distribution *a priori* des paramètres et de la vraisemblance *a posteriori* pour ce sujet.

Du fait de la non linéarité du modèle, le profil "moyen" obtenu avec la fonction f en les effets fixes μ diffère de la moyenne des observations prédites à chaque temps par la fonction f en les paramètres individuels, comme illustré sur la figure 1.6.

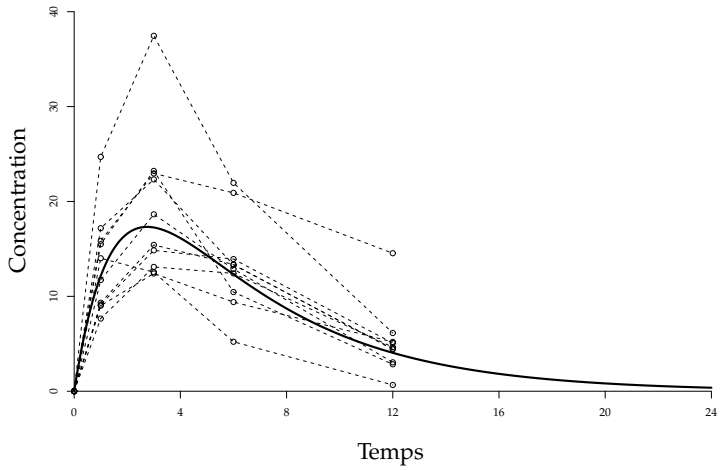


FIG. 1.6 – Profil "moyen" pour un modèle pharmacocinétique à un compartiment d'absorption et d'élimination d'ordre 1 (trait continu) superposé à différents profils individuels obtenus par simulation avec des variabilités inter-sujets et résiduelle de respectivement 30% et 20% (traits pointillés).

1.5.2 Les méthodes d'estimation

Dans les MNLEM, la vraisemblance n'a pas de solution analytique. Des algorithmes d'estimation par maximum de vraisemblance spécifiques ont donc été développés. Les premiers algorithmes étaient basés sur une linéarisation du modèle au premier ordre autour des effets fixes comme dans l'algorithme *First-order* (FO, Sheiner et al. (1972)) ou autour des effets aléatoires comme dans l'algorithme *First-Order Conditional Estimation* (FOCE, Lindstrom et Bates (1990)) ce qui permet de se ramener à un modèle linéaire mixte dont la fonction de vraisemblance a une forme explicite. Ces algorithmes ont tendance à produire des estimateurs biaisés et peu précis en présence de forte variabilité individuelle et/ou lorsque le nombre d'observations par sujet croît moins vite que le nombre de sujets (Ge et al., 2004; Vonesh, 1996). Wolfinger (1993) a par la suite proposé une méthode d'estimation où la vraisemblance est calculée par une approximation de Laplace. Ces méthodes sont mises en œuvre dans le logiciel NONMEM (Sheiner et Beal, 1998), très largement utilisé dans l'industrie pharmaceutique mais aussi dans le milieu de la recherche. Les méthodes FO et l'approximation de Laplace sont aussi mises en œuvre dans le logiciel SAS (SAS Institute Inc, 2004). Par ailleurs, l'algorithme FOCE et l'approximation de Laplace sont mis en œuvre respectivement dans les fonctions nlme et nlmer du logiciel R (R Development Core Team, 2008).

L'amélioration de la puissance de calcul des ordinateurs a permis de considérer des méthodes d'estimation plus précises mais aussi plus consommatrices en temps de calcul. Ainsi, la quadrature de Gauss adaptative est un développement de l'approximation de Laplace à un ordre plus élevé dont les noeuds et les poids sont déterminés pour mieux couvrir la distribution des effets aléatoires (Pinheiro et Bates, 2000). Cette méthode, mis en œuvre dans le logiciel SAS, est plus précise et moins biaisée mais nécessite des temps de calculs plus longs et peut se montrer moins stable. Les travaux de Commenges et al. (2005) ont permis de surmonter les problèmes d'instabilité liés aux calculs des dérivées première et seconde de la fonction de vraisemblance, mais ces développements ne sont pas disponibles dans un logiciel aujourd'hui.

Plus récemment, l'algorithme *Stochastic Approximation Expectation Maximisation* (SAEM) a été développé par Kuhn et Lavielle (2005). Cette méthode est une version stochastique de l'algorithme EM où les paramètres individuels sont considérés comme les valeurs manquantes. L'étape d'estimation se décompose en une simulation de paramètres individuels avec une méthode de Monte Carlo par chaînes de Markov, suivie du calcul de statistiques suffisantes de la vraisemblance des données complètes. La maximisation de ces statistiques suffisantes permet d'obtenir de nouvelles valeurs des paramètres, réutilisés à la première étape de l'itération suivante jusqu'à convergence de l'algorithme. L'algorithme SAEM est implémenté dans le logiciel MONOLIX (Lavielle, 2008) qui est développé par le Pr. Marc Lavielle.

1.5.3 Les tests et méthodes de sélection de modèles

Les deux méthodes de tests principalement utilisées pour détecter un effet gène sur un paramètre pharmacocinétique dans les MNLEM sont :

i. l'analyse de variance à un facteur (ANOVA) où les estimations Bayesiennes empiriques (EBE) des paramètres individuels calculées à partir du modèle sans covariable sont comparées entre les différents groupes de génotypes. Cette méthode est surtout employée pour les analyses exploratoires sur les covariables comme dans le travail de Hirt et al. (2008)

ii. le test du rapport de vraisemblance (LRT) qui permet de comparer les vraisemblances des modèles avec et sans la covariable. Ce test est surtout utilisé dans les stratégies de sélection ascendante du modèle de covariables comme dans l'analyse de Yamasaki et al. (2008)

Le test de Wald est une alternative au LRT moins utilisée en pharmacocinétique de population qui repose sur les erreurs d'estimation et permet d'évaluer la significativité des coefficients d'effet estimés par le modèle incluant la covariable génétique.

Plusieurs travaux ont évalué l'erreur de type I et la puissance de ces tests dans le contexte des études de pharmacocinétique de population. La majorité de ces travaux a utilisé des algorithmes basés sur une linéarisation du modèle. Ainsi, Wählby et al. (2001); White et al. (1992); Bonate (2005); Gobburu et Lawrence (2002); Lee (2001); Panhard et Mentré (2005) et Panhard et al. (2007) ont montré qu'avec FO ou FOCE augmenter le nombre de prélèvements entraîne une inflation de l'erreur de type I du LRT et du test de Wald. Plus récemment, Panhard et Samson (2009) ont montré que cette relation n'est pas observée avec l'algorithme SAEM. Dans le cadre des MNLEM, les propriétés statistiques du LRT et du test de Wald, qui sont dérivés directement de la théorie de la vraisemblance, ne sont conservées qu'asymptotiquement. Dans ce sens, Panhard et Mentré (2005) et Wählby et al. (2001) ont montré que l'erreur de type I du LRT et du test de Wald est augmentée significativement lorsque N est petit. En ce qui concerne les tests sur les EBE, Bonate (2005) et Comets et Mentré (2001) n'ont pas observé d'inflation de l'erreur de type I sur les plans d'expérience classiques, alors que dans les essais en cross-over Panhard et Mentré (2005) ont montré une inflation de l'erreur de type I de ces tests lorsque le nombre de prélèvements est réduit. Toutes ces études sont résumées et ordonnées par date de publication dans une table, en discussion de l'article présenté dans la partie 2.2. Pour la plupart, ces études n'ont considéré que des covariables continues ou binaires, cas classiques en pharmacologie clinique.

Depuis ces 10 dernières années, des critères de sélection qui ont été développés comme une alternative aux tests, sont de plus en plus utilisés en pharmacocinétique/pharmacodynamie dans le cadre de la sélection de modèles. Ces critères reposent sur un principe simple : l'augmentation de la vraisemblance associée

à l'ajout d'un paramètre est pondérée par une pénalisation prenant en compte le nombre de paramètres dans le modèle et/ou le nombre de données à disposition. Ils permettent ainsi de comparer des modèles non-emboîtés. Si le critère d'Akaike (AIC, Akaike (1974)) a été développé à partir de la théorie de l'information et de la distance de Kullback-Leibler (Kullback et Leibler, 1951), le second critère le plus utilisé, le critère d'information de Bayes (BIC, Schwartz (1978)), est lui issu de la théorie Bayésienne et quantifie l'incertitude attachée à chaque modèle. D'autres critères dérivés de l'AIC ont été développés, pour les études avec des effectifs réduits (Sugiura, 1978) ou encore pour améliorer leurs propriétés asymptotiques (Bozdogan, 1987). Les propriétés de ces critères de sélection ont peu été explorées dans le cadre des MNLEM mais plutôt dans les modèles linéaires généralisés (Burnham et Anderson, 2002), mettant en évidence une tendance à la sélection de modèles sur-paramétrés avec l'AIC et sous-paramétrés avec le BIC.

1.5.4 Les plans d'expériences

Si un nombre important de prélèvements chez un volontaire sain est peu problématique, diminuer ce nombre devient nécessaire chez le patient. Notamment dans les populations sensibles comme les enfants et les personnes âgées, où toute intervention invasive doit être limitée. Les autorités de santé ont statué récemment sur la nécessité d'inclure dans les études pharmacogénétiques un nombre suffisant de patients pour avoir une bonne représentation des différents génotypes (Committee for human medicinal products, 2007). Toutefois, pour des raisons économiques et de pratique clinique, plus le nombre de patients augmente plus le nombre de prélèvements par patient est amené à diminuer.

Six des études publiées entre 2008 et septembre 2009 incluaient entre 50 et 100 patients avec une cinétique de plusieurs prélèvements pour chaque patient. Deux études incluaient au moins 70 patients avec des concentrations résiduelles et un sous-groupe avec une cinétique de plusieurs prélèvements. Quatre autres études incluaient de 70 à 130 patients avec seulement 1 à 2 concentrations. Yang et al. (2009) ont réparti leurs 48 patients dans plusieurs groupes de temps de prélèvements.

Il existe donc une grande diversité dans les études pharmacogénétiques réalisées actuellement en terme de plan d'expérience. Or, le choix du plan d'expérience agit de façon déterminante sur la puissance des tests qui dans les MNLEM dépend du nombre de patients ainsi que du nombre et des temps de prélèvements (Retout et al., 2007b).

1.6 Objectifs de la thèse

Du fait de la multiplicité des génotypes et de leur représentation déséquilibrée dans la population générale, nous avons voulu évaluer les propriétés statistiques de l'ANOVA, du LRT et du test de Wald pour détecter un effet gène sur un paramètre pharmacocinétique

dans le cadre des MNLEM. De la même manière, nous avons voulu explorer les propriétés des critères AIC, BIC et leurs dérivés. Nos travaux méthodologiques ont été réalisés par l'intermédiaire d'études de simulation. Dans ces simulations, la covariable génétique est entrée dans le modèle de façon linéaire.

Les résultats de l'évaluation des tests et des critères de sélection sont présentés dans la partie 2.1 avec les algorithmes FO et FOCE, et pour l'évaluation des tests uniquement dans la partie 2.2 avec l'algorithme SAEM. Nous avons aussi réalisé une évaluation de l'impact du plan d'expérience sur les performances des tests où nous avons considéré deux protocoles pragmatiques et un protocole optimisé. Ce travail est présenté dans la partie 2.2. Ces différents résultats nous ont amené à proposer des alternatives aux tests asymptotiques dont l'évaluation est présentée dans la partie 2.3.

Enfin, nous avons appliqué les conclusions de nos travaux méthodologiques à l'analyse de trois études pharmacogénétiques présentées au chapitre 3. La première étude présentée dans la partie 3.1 porte sur l'influence de cinq polymorphismes génétiques sur la pharmacocinétique ainsi que sur l'effet et la toxicité à court terme de l'indinavir dans l'essai COPHAR2-ANRS 111 ayant pour but d'évaluer l'intérêt de l'adaptation précoce de dose chez des patients naïfs de traitement par inhibiteur de protéase. La deuxième étude décrite dans la partie 3.2 explore l'influence du polymorphisme des gènes *CYP* 2D6 et *CYP* 2C9 sur la pharmacocinétique du produit parent et du métabolite actif d'un anti-psychotique en développement dans les laboratoires pharmaceutiques Servier. Enfin la troisième étude présentée dans la partie 3.3 traite de l'influence de quatre polymorphismes sur la pharmacocinétique de la névirapine chez des patients cambodgiens naïfs de traitement inclus dans l'essai PECAN-ANRS 12414.

Chapitre 2

Travaux méthodologiques

2.1 Tests et méthodes de sélection de modèles pour détecter un effet gène sur un paramètre pharmacocinétique

2.1.1 Résumé

La pharmacogénétique étudie comment les polymorphismes génétiques codants pour des protéines de transport ou des enzymes du métabolisme peuvent expliquer la variabilité pharmacocinétique d'un médicament. L'analyse de variance (ANOVA) et le test du rapport de vraisemblance (LRT) sont couramment utilisés pour explorer l'influence des polymorphismes génétiques dans les études pharmacocinétiques qui utilisent les modèles non linéaires à effets mixtes (MNLEM). L'ANOVA est principalement utilisée pour les analyses exploratoires et le LRT pour l'inclusion des covariables génétiques dans le modèle. Le test de Wald qui est une alternative au LRT reposant sur les erreurs d'estimation, et les critères de sélection tels que le critère d'Akaike (AIC) ou le critère d'information de Bayes (BIC) sont relativement peu utilisés dans ce domaine.

Dans cette première étude, nous avons évalué par simulation les performances de l'ANOVA, du LRT, du test de Wald et des critères de sélection pour détecter l'influence d'un effet gène sur un paramètre pharmacocinétique. Le cadre pharmacocinétique de cette étude de simulation est inspiré de la sous-étude réalisée sur l'indinavir dans l'essai COPHAR 2-ANRS 111 (Duval et al., 2009; Bertrand et al., 2009) et décrite dans la partie 3.1. Les valeurs des effets fixes et variances utilisées pour générer les paramètres individuels sont issues d'une analyse préliminaire des données. La distribution et l'effet associés aux polymorphismes génétiques simulés sont ceux des exons 26 et 21 du gène ABCB1 codant pour la P-glycoprotéine tels que rapportés dans la littérature (Sakaeda et al., 2002). Nous avons simulé les concentrations de N=40 sujets avec n=4 temps d'observations à 1, 3, 6 et 12h après administration de la dose d'indinavir sous l'hypothèse nulle (H_0) et sous l'hypothèse alternative d'un effet génétique (H_1). Un second plan d'expérience a été simulé sous H_0 incluant N=200 patients pour se rapprocher de

conditions asymptotiques.

Nous avons évalué l'erreur de type I (sous H_0) et la puissance (sous H_1) des tests, ainsi que la capacité à choisir le modèle simulé de huit stratégies de sélection de modèles. Les trois premières stratégies sont basées sur les tests, les autres stratégies reposent sur 5 critères de sélections où, suivant le principe de parcimonie, le modèle avec le critère le plus faible est sélectionné. Tout d'abord, nous avons considéré le critère d'Akaike ainsi que deux alternatives développées respectivement dans le cadre d'effectifs restreints (AIC_C, Sugiura (1978)) et pour améliorer la consistance de l'Akaike ($CAIC$, Bozdogan (1987)) :

$$AIC = -2L + 2P_{pop}$$
$$AIC_C = -2L + \frac{2P_{pop}(P_{pop}+1)}{n_{tot} - P_{pop} - 1}$$
$$CAIC = -2L + 2P_{pop}log(n_{tot}+1)$$

Les deux derniers critères évalués sont dérivés du facteur de Bayes qui est une mesure de l'incertitude associée au modèle sélectionné, le BIC et une alternative développée dans le cas de données répétées le BIC_C (Raftery, 1995) :

$$BIC = -2L + P_{pop}log(n_{tot})$$
$$BIC_C = -2L + P_{pop}log(N)$$

Dans ces formules, L est la log-vraisemblance du modèle, P_{pop} le nombre de paramètres estimés, N le nombre de sujets et n_{tot} le nombre total d'observations. Afin de prendre en compte l'impact de l'algorithme d'estimation, les données ont été analysées avec les méthodes FO et FOCE dans le logiciel NONMEM version V. Enfin, ces différentes méthodes ont été appliquées à l'analyse de l'influence du polymorphisme des exons 26 et 21 du gène ABCB1 sur les concentrations d'indinavir recueillies dans l'essai COPHAR 2.

Nous avons rencontré des difficultés en termes de convergence et de calcul de la matrice de variance d'estimation avec ces algorithmes, de façon plus prononcée avec FOCE. Seule l'ANOVA conserve une erreur de type I non significativement différente de 5% avec les deux algorithmes d'estimation. Le test de Wald et le LRT ont une erreur de type I significativement plus élevée. Cette inflation particulièrement importante avec FO, au dessus de 40% pour le LRT, est plus raisonable avec FOCE, autour de 10%, et disparaît sur le plan d'expérience avec 200 sujets pour cet algorithme. Avec FOCE, les puissances corrigées de l'ANOVA et du LRT sont comparables et autour de 70% alors que la puissance corrigée du test de Wald est en dessous de 25%. La faible puissance de ce test résulte d'une corrélation observée entre les coefficients d'effet et leur erreur d'estimation qui pourrait provenir de la mise en œuvre du calcul de la matrice de variance d'estimation pour la version du logiciel NONMEM utilisée. Les résultats obtenus par les stratégies de sélection basées sur les tests sont globalement similaires aux résultats des tests dont elles sont issues. Pour les stratégies basées sur les critères, nous confirmons la tendance

systématique de l'AIC et de l'AIC_C à la sélection de modèles sur-paramétrés alors que le BIC et le $CAIC$ sélectionnent le modèle simulé à 90% sous H_0. Sur les données de l'essai COPHAR2, nous ne mettons pas en évidence d'effet des polymorphismes du gène ABCB1 sur la pharmacocinétique de l'indinavir.

Avec les algorithmes FO et FOCE, l'erreur de type I des tests asymptotiques est augmentée significativement. Si le BIC et le $CAIC$ obtiennent des performances satisfaisantes, les critères AIC et AIC_C ne devraient pas être utilisés pour la construction de modèle de covariables.

Cette étude a fait l'objet d'un article publié dans la revue *Journal of Biopharmaceutical Statistics*.

2.1.2 Article 1 (publié)

Journal of Biopharmaceutical Statistics, 18: 1084–1102, 2008

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543400802369012

Taylor & Francis
Taylor & Francis Group

COMPARISON OF MODEL-BASED TESTS AND SELECTION STRATEGIES TO DETECT GENETIC POLYMORPHISMS INFLUENCING PHARMACOKINETIC PARAMETERS

Julie Bertrand, Emmanuelle Comets, and France Mentré
UFR de Médecine–Site Bichat, UMR 738 INSERM Paris Diderot, Paris, France

We evaluate by simulation three model-based methods to test the influence of a single nucleotide polymorphism on a pharmacokinetic parameter of a drug: analysis of variance (ANOVA) on the empirical Bayes estimates of the individual parameters, likelihood ratio test between models with and without genetic covariate, and Wald tests on the parameters of the model with covariate. Analyses are performed using the FO and FOCE method implemented in the NONMEM software. We compare several approaches for model selection based on tests and global criteria. We illustrate the results with pharmacokinetic data on indinavir from HIV-positive patients included in COPHAR 2-ANRS 111 to study the gene effect prospectively. Only the tests based on the EBE obtain an empirical type I error close to the expected 5%. The approximation made with the FO algorithm results in a significant inflation of the type I error of the LRT and Wald tests.

Key Words: Genetics; Model selection; NONMEM; Nonlinear mixed effects model; Pharmacokinetics; Test.

1. INTRODUCTION

Pharmacokinetics studies the time course of a drug in the body (Gabrielsson and Weiner, 1999). The variability in the pharmacokinetics of drugs when administered to different subjects is often important and should be studied to improve the use of drugs, avoid toxic events, and allow individualization of therapy. The contribution of genetic factors to this variability is potentially important (Licinio and Wong, 2002). Some genes have already been the subject of much attention, for example the ABCB1 gene coding for the P glycoprotein (P-gP) found on the main exchange barriers (Marzolini et al., 2003). The involvement of this protein in drug absorption processes has been demonstrated directly in animals and indirectly in humans. For example, co-administration of paclitaxel and cyclosporine, respectively a substrate and an inhibitor of P-gP, has been shown to increase the bioavailability of paclitaxel (Meerum et al., 1999). More recently, Yamaguchi and colleagues demonstrated the impact of ABCB1 polymorphism on

Received July 26, 2007; Accepted November 28, 2007
Address correspondence to Julie Bertrand, UFR de Médecine–Site Bichat, UMR 738 INSERM Paris Diderot, 16 rue Henri Huchard, Paris 75018, France; E-mail: julie.bertrand@bichat.inserm.fr

paclitaxel pharmacokinetics in patients with ovarian cancer (Yamaguchi et al., 2006). The ABCB1 gene is composed of 209 kb (Bodor et al., 2005), and to date 28 polymorphisms concerning a nucleotide modification have been described, including several with potential clinical impact (Sakaeda et al., 2002). In the present study, we consider data from patients treated with indinavir in the COPHAR2-ANRS 111 study investigating the benefit of early therapeutic drug monitoring in anti-retroviral therapy. Since the ABCB1 genetic polymorphism was found to have an influence on the pharmacokinetics of protease inhibitors (Fellay et al., 2002), patients were genotyped for two exons (exon 21 and 26) of the ABCB1 gene.

The influence of genetic polymorphism on concentration data is usually analyzed using noncompartmental analysis (NCA). The area under the concentration vs. time curve as well as model-independent parameters are calculated for each individual concentration profile using, for example, the log-trapezoidal method (Gabrielsson and Weiner, 1999) and individual pharmacokinetic (PK) parameters are then compared between the different genotype groups using analysis of variance (ANOVA) (Inomata et al., 2005). More sophisticated approaches using nonlinear PK models in individual regression (Min et al., 2004) or mixed effects models (Taguchi et al., 2006) have also been applied to genetic data in PK studies. Nonlinear mixed effects models (NLMEM) require fewer blood samples in each patient and can be important in special populations such as patients with acute disease or neonates, for whom extensive sampling is obviously impractical.

Various methods can be used to include pharmacogenetic information in NLMEM (Comets et al., 2007). Empirical Bayes estimates (EBE) of the individual parameters can be computed from the fit of a model with no covariate and compared between the different genotype groups with an ANOVA to test the genetic polymorphism effect (Henningsson et al., 2005). Another approach is to perform Wald tests on the estimates of the gene effect coefficients in the covariate model (Kerbusch et al., 2003), whereas stepwise model building is frequently based on the comparison of models with and without gene effect using the likelihood ratio test (LRT) (Mamiya et al., 2000). The first purpose of this work is to assess the statistical properties of these three different strategies to test for a gene effect through a simulation study. The setting for the simulation study is based on the COPHAR2-ANRS 111 study. We have also taken the impact of estimation methods into account, comparing results from the FO and FOCE methods implemented in the NONMEM (nonlinear mixed effect model) software (Sheiner and Beal, 1998). The type I error of the tests is evaluated using simulations under the null hypothesis H_0 of no gene effect for two designs (40 patients as in the study, 200 patients to examine the influence of sample size), whereas the power is compared using simulations under an alternative hypothesis H_1.

In the same simulation study, we also examine model selection strategies, in which the aim is to choose the best covariate model for the gene effect. We compare test-based strategies with general selection criteria that have been proposed, such as the Bayes information criterion (BIC) (Schwartz, 1978) and the Akaike information criterion (AIC) (Akaike, 1974). According to the parsimony principle, these criteria penalize the log-likelihood by the number of model parameters, which limits overfitting. Alternative criteria derived from the AIC have been developed in order to deal with small samples or to improve AIC asymptotic consistency. The properties of these criteria have been studied in generalized linear models (Burnham

and Anderson, 2002), such as the trend to select an over-parameterized model with AIC and under-parameterized model with BIC. As well, AIC and BIC have been widely used in population PK and PK/PD studies during the past decade. In the present study, we compare through a simulation study their properties for nonlinear mixed effects model selection.

We first introduce the model, the notations, the different tests and model building strategies under study, and the estimation methods. Then we describe the case study and the simulation study based on its design. Next we present the results of the simulation study and the application to the real data. Finally we discuss our findings.

METHODS

Model and Notations

Let the function f denote the PK model, which depends nonlinearly on its parameters θ. The concentration $y_{i,j}$ at time $t_{i,j}$ for subject $I = 1, \ldots, N$ and measurement $j = 1, \ldots, n_i$ is given by:

$$y_{i,j} = f(t_{i,j}, \theta_i) + \varepsilon_{i,j} \tag{1}$$

where θ_i is the vector of P PK parameters for the ith individual and $\varepsilon_{i,j}$ the residual error, assumed to be normally distributed with zero mean and variance $\sigma^2_{i,j}$. Here we assume a proportional error model:

$$\sigma^2_{i,j} = \sigma^2 f(t_{i,j}, \theta_i)^2 \tag{2}$$

We assume a multivariate log-normal distribution for the vector of individual parameters; θ_i is then expressed as:

$$\theta_i = \mu e^{bi} \tag{3}$$

where μ is the vector of fixed effects and b_i the vector of random effects. The random effects b_i are assumed to be independent of $\varepsilon_{i,j}$ and normally distributed with zero mean and variance matrix Ω. In this work, we used a diagonal Ω with interindividual variance of the pth parameter ω^2_p, however an unspecified positive definite matrix can be assumed.

For simplicity, let us assume that our aim is to detect the effect of a single nucleotide polymorphism (SNP) on one PK parameter, for instance the pth component of θ; θ^p. Let C denote the wild-type allele and T the mutant. An individual can be one of three genotypes (e.g., CC for the wild-type homozygotes, CT or TT). Let G_i denote the genotype for subject i, and let $\beta(G_i) = \beta_0$, β_1, or β_2 for $G_i = \text{CC}$, CT, or TT, respectively. We write the model for the genetic polymorphism effect in subject i as:

$$\theta^p_i = \mu^p \beta(G_i) e^{b_i p} \tag{4}$$

We assume that CC is the reference class so that $\beta_0 = 1$.

Let model M_{base} be the model in the absence of genetic polymorphism effect: $\{\beta_0 = \beta_1 = \beta_2 = 1\}$, (CC = CT = TT). There are three models including the

gene effect as a covariate: a complete model, where the mean of the parameter is different in the three groups, $M_{\text{mult}} : \{\beta_0 = 1, \beta_1 \neq \beta_2 \neq 1\}$, (CC $\neq$ CT $\neq$ TT), and two reduced models, $M_{\text{recessive}} : \{\beta_0 = \beta_1 = 1, \beta_2 \neq 1\}$, (CC = CT $\neq$ TT) and $M_{\text{dominant}} : \{\beta_0 = 1, \beta_1 = \beta_2 \neq 1\}$, (CC $\neq$ CT = TT).

In the following, L corresponds to the model log-likelihood, P_{pop} represents the number of population model parameters (mean PK parameters, covariate coefficients, variances and error model parameters), N is the sample size, and n_{tot} is the total number of observations. The standard error and covariance of the parameter estimates are abbreviated by SE and cov, respectively.

Tests for Genetic Polymorphism Effect

In this section, we examine different tests based on NLMEM that can be used to test the existence of a genetic polymorphism effect on one parameter of a PK population model.

Analysis of variance (ANOVA). Data are analysed with a model with no covariates (M_{base}) and empirical Bayes estimates (EBE) of the individual PK parameters are computed. A one-way analysis of variance is used to detect differences in these EBE between the different genetic groups.

Wald test. The data are analyzed using the complete model M_{mult}, and the significance of the parameters is assessed using a global Wald test:

$$W = \begin{pmatrix} \beta_1 - 1 \\ \beta_2 - 1 \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} \beta_1 - 1 \\ \beta_2 - 1 \end{pmatrix} \tag{5}$$

where Σ represents the variance covariance matrix of parameters β_1 and β_2. The statistic W is compared to the critical value of a χ^2 with two degrees of freedom.

Likelihood ratio test (LRT). The third test relies on the comparison between the model with no covariate effect M_{base} and the complete model M_{mult}. These two nested models are compared with the LRT. The test statistic $S_{LR} = -2(L_{\text{base}} - L_{\text{mult}})$ is compared with a χ^2 with two degrees of freedom, where L_{base} and L_{mult} are the log-likelihood of M_{base} and M_{mult}, respectively. The two degrees of freedom correspond to the difference in the number of population parameters between the two models.

Strategies for Model Building

In this section, we examine a second aspect, which is model selection. The different strategies under study provide decision rules for covariate inclusion in order to get the best model. Figure 1 represents in a diagram the decision path of the three strategies using tests. When performing multiple tests, a correction that preserves the false discovery rate is used (Benjamini and Hochberg, 1995).

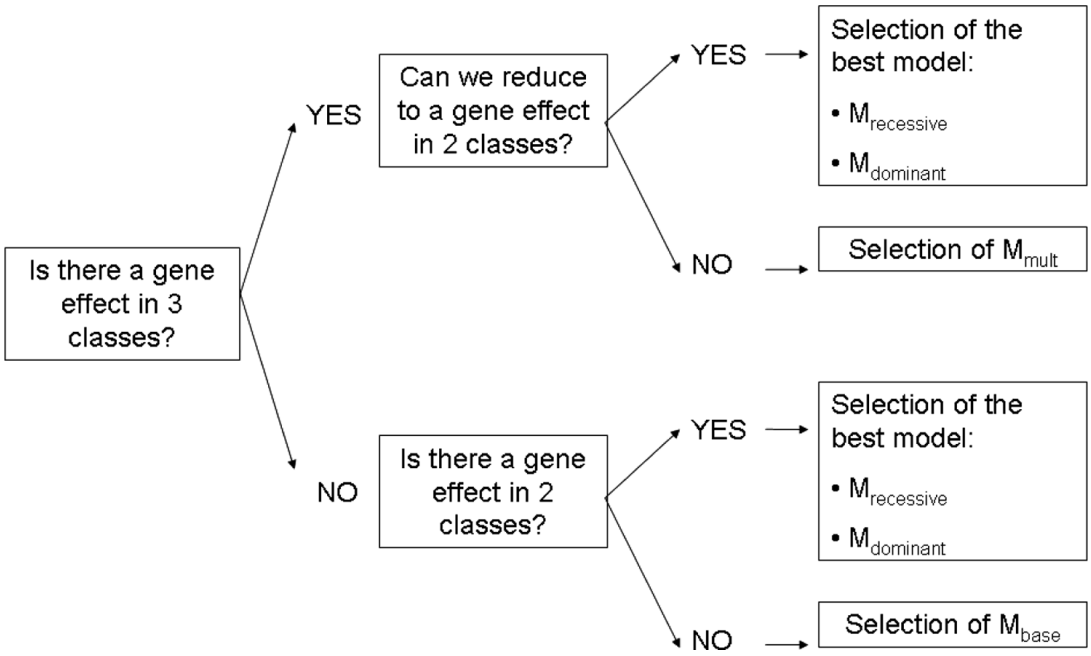


Figure 1 Decision path used to choose the best model for the selection strategies based on tests.

Selection based on EBE. If the test that compares means of EBE obtained with M_{base} between CC, CT, and TT using an ANOVA is significant, the means of the EBE obtained with M_{base} are compared between the CC and the CT, on one hand, and between the CT and the TT, on the other hand, using t-tests. If none or one test is significant, the one with the lower p-value leads to model selection, or else M_{mult} is selected. For example, if the test comparing the CC to the CT has the lower p-value, M_{dominant} is chosen.

If the global test is nonsignificant, the means of the EBE obtained with M_{base} are compared using t-tests opposing: (i) the CC to the CT and TT put together, and (ii) the TT to the CT and CC put together. If one or both tests are significant, the model corresponding to the test with the lower p-value is selected, else M_{base} is selected. For example if the test comparing the CC to the group formed by the CT and the TT has the lower p-value, M_{dominant} is chosen.

Selection based on Wald test. If the global Wald test comparing W to a χ^2 with two degrees of freedom is significant, two Wald tests on coefficients estimated in M_{mult} are then realized, comparing both $\frac{(\beta_1-1)^2}{SE^2(\beta_1)}$ and $\frac{(\beta_1-\beta_2)^2}{SE^2(\beta_1)+SE^2(\beta_2)-2\text{cov}(\beta_1,\beta_2)}$ with a χ^2 with one degree of freedom. If none or one test is significant, the one with the lower p-value leads to model selection, or else M_{mult} is kept. As an example, if $\frac{(\beta_1-1)^2}{SE^2(\beta_1)}$ has the lower p-value, M_{dominant} is chosen.

If the global test is nonsignificant, data are analyzed using the two models with the gene effect in two classes and Wald tests are performed on the coefficients estimated in $M_{\text{recessive}}$ and M_{dominant}. The two statistics $\frac{(\beta_{\text{rec}}-1)^2}{SE^2(\beta_{\text{rec}})}$ and $\frac{(\beta_{\text{dom}}-1)^2}{SE^2(\beta_{\text{dom}})}$ are compared with a χ^2 with one degree of freedom. If one or both tests are significant, the model is chosen based on the test with the lower p-value, or else M_{base} is conserved. By way of example, if $\frac{(\beta_{\text{dom}}-1)^2}{SE^2(\beta_{\text{dom}})}$ has the lower p-value, M_{dominant} is selected.

Selection based on LRT. If the LRT comparing M_{base} with M_{mult} is significant, then two LRT are performed comparing M_{mult} with the nested models $M_{dominant}$ and $M_{recessive}$. If none or one test is significant, the test with the higher p-value leads the model choice, or else M_{mult} is selected. As an example, if the test comparing M_{mult} with $M_{dominant}$ obtains the higher p-value, $M_{dominant}$ is selected.

If the global test is nonsignificant, M_{base} is compared with $M_{dominant}$ and $M_{recessive}$ using LRT. Then, if one or both tests are significant, the model is chosen based on the test with the lower p-value, or else M_{base} is selected. For example, if the test comparing M_{base} to $M_{dominant}$ obtains the lower p-value, $M_{dominant}$ is selected.

Selection based on information criteria. As an alternative, criteria such as AIC or BIC can be used for model selection. Here, the model with the lowest criterion is chosen. Both of these criteria balance the log-likelihood with the number of parameters in the population model. According the parsimony principle, a simpler model is preferred for equivalent information gain. The AIC is written as

$$\mathrm{AIC} = -2L + 2P_{pop} \tag{6}$$

The expression of the BIC involves the total number of observations:

$$\mathrm{BIC} = -2L + P_{pop}\log(n_{tot}) \tag{7}$$

The BIC is related to the Bayes factor, which is a measure of the strength of evidence in favor of a given model and can thus be used to quantify model uncertainty.

Other criteria have been defined, derived from those previously shown: the corrected AIC (AICc) defined by Akaike for small samples, i.e., when $n_{tot}/P_{pop} < 40$ (Sugiura, 1978), and the consistent AIC (CAIC) (Bozdogan, 1987). They are derived from AIC and involve the total number of observations:

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2P_{pop}(P_{pop}+1)}{n_{tot} - P_{pop} - 1} \tag{8}$$

$$\mathrm{CAIC} = -2L + P_{pop}(\log(n_{tot}) + 1) \tag{9}$$

Another formulation for BIC has also been proposed, BICc, where n_{tot} is replaced by the number of subjects (Raftery, 1995):

$$\mathrm{BICc} = -2L + P_{pop}\log(N) \tag{10}$$

Estimation Methods

In NLMEM, the parameters and their standard error are mainly estimated using maximum likelihood. However the likelihood function for these models is expressed as an integral and has no analytical solution. Specific algorithms have therefore been proposed to perform the maximization. The approach most frequently used relies on first-order approximations of the likelihood function. This approach has been implemented in the NONMEM software version V, which is the most frequently used software in PK/PD analyses. In NONMEM two approximations can be used: the first-order method (FO), relying on a first-order

linearization of the likelihood function around $b_i = 0$, and the first-order conditional estimation method (FOCE), relying on a first-order linearization of the likelihood function around the estimates of the individual random effects. Although the FO method suffers from bias and lack of precision it is still frequently used because it is faster and has less convergence problems than FOCE. In this paper, we compare both estimation methods. More precisely, we use the FOCE with interaction method, allowing possible interaction between inter- and intravariability.

Real Data and Simulation Study

Real data. We illustrate the different approaches on data from a PK sub-study of the COPHAR 2-ANRS 111 study, a multicenter noncomparative pilot trial of early therapeutic drug monitoring in HIV-positive patients naïve of treatment. The objective of the trial was to assess the benefit of a pharmacological intervention after measurement of trough plasma concentrations of protease inhibitors (Mentré et al., 2005). We focus on the PK sub-study from the group of patients receiving indinavir boosted with ritonavir. Patients were genotyped for the ABCB1 exons 21 and 26 to investigate genetic polymorphism impact on pharmacokinetics of the protease inhibitors, which are well-known substrates of the P-gp (Fellay et al., 2002).

Forty-two patients were included, one patient withdrew from the study, and one switched to another protease inhibitor during the first week of treatment. We therefore obtained PK data for 40 patients (27 men, 13 women) with an average age of 36.5 years. The PK profiles were determined at 1, 3, 6, and 12 h after administration of the drug at a date 2 weeks after the treatment onset. One patient had missing information for the two genotypes, and ABCB1 exon 26 genotype was missing in two other patients.

Simulated data. The design used in the simulation mimics that of the application data set. We simulate PK studies of $N = 40$ (equivalent to the application sample size) and $N = 200$ patients with four samples (1, 3, 6, and 12 h after dose) at steady state. The same b.i.d. doses of 400 mg for indinavir and 100 mg for ritonavir are assumed for all patients. The concentrations are simulated using the steady-state one-compartment model with first-order absorption and elimination that was used to model the indinavir concentrations in the COPHAR 1-ANRS 102 study (Brendel et al., 2005):

$$f(\theta, t) = \frac{D}{V/F}\frac{k_a}{k - k_a}\left(\frac{\exp(-k_a t)}{1 - \exp(-k_a \tau)} - \frac{\exp(-kt)}{1 - \exp(-k\tau)}\right) \tag{11}$$

where F represents the bioavailability, k_a the absorption rate, k the elimination rate, V the volume of distribution, and $\tau = 12$ h the time between two doses. We use the following population parameters: $V/F = 102$ L with interindividual variability $\omega_V = 41.3\%$, $k_a = 1.4\,\text{h}^{-1}$ with $\omega_{k_a} = 113\%$, $k = 0.2\,\text{h}^{-1}$ with $\omega_k = 26.4\%$ obtained with a preliminary analysis of the real data using the FO method in NONMEM, and a residual error of 20%. The measurements below the quantification limit (BQL) are treated in the analysis using a standard approach: the first value in a series of BQL was set to LOQ/2 and the remaining values were discarded (Beal, 2005).

Under H_0, 1,000 data sets are simulated with a design of $N = 40$ patients and 1,000 with a design of $N = 200$ patients. We simulate a combination of SNP on two exons located on the same chromosome. The bioavailability F is assumed to depend on the diplotype. Because the model is parameterized as V/F, k_a, and k, this is equivalent to assuming that V/F depends on the diplotype. The distribution of the exons mimics that of exon 26 and exon 21 of the ABCB1 gene as reported by Sakaeda and colleagues (Sakaeda et al., 2002)—we note C and G, respectively, the wild-type allele for the 2 exons and T the mutant allele. With those properties, for data sets with 40 patients we expect on average 9 individuals with a CC genotype, 18 with CT genotype, and 13 with TT genotype for exon 26.

Under the alternative hypothesis, we assume the following effect of the two polymorphisms (modified from Equation (4)), where G_{26i} denotes the genotype for the exon 26 and G_{21i} the genotype for the exon 21:

$$(V/F)_i = V/F\ \beta(G_{26i})\delta(G_{21i})e^{b_i} \tag{12}$$

where $\beta(G_{26i})$ is β_0, β_1, or β_2 if $G_{26i} = \text{CC}$, CT, or TT as previously, and $\delta(G_{21i})$ is δ_0, δ_1, or δ_2 if $G_{21i} = \text{GG}$, GT, or TT. Under H_1 we set $\beta(G_{26}) = 1$, 1.2, and 1.6, and $\delta(G_{21}) = 1$, 1.1, and 1.3. These values were chosen to provide a good power for the detection of one SNP effect in the context of the simulations while remaining consistent with results found in the literature concerning the effect of ABCB1 polymorphism. With those genetic coefficient effects and the distribution from the literature, we simulated 100 data sets with the $N = 40$ design. When we computed the $(V/F)_i$ EBE from M_{base} and performed bilateral t-tests to compare the wild homozygotes and the mutant homozygotes for the exon 26, we obtained a power of 80% (Machin et al., 1997).

The simulation of each data set is performed as follows. The set of possible genotypes for exon 26 and exon 21 is $S = \{$CC-GG, CC-GT, CC-TT, CT-GG, CT-GT, CT-TT, TT-GG, TT-GT, TT-TT$\}$ with corresponding simulated frequencies $f = \{0.2, 0.02, 0.02, 0.05, 0.38, 0.05, 0.04, 0.04, 0.2\}$. For each individual the genotypes are drawn from this distribution. Under H_1, both genotypes condition the value of the fixed effects for V/F according to Equation (11). Then we simulate a random effect vector b_i from a normal distribution $N(0, \Omega)$, yielding the individual parameter vector θ_i according to Equation 3. The concentrations are computed using these parameters. Finally we add a residual error, generated from a normal distribution $N(0, \sigma^2 f(\theta, t_{i,j}))$, to each predicted concentration to obtain the simulated concentration.

Simulations were performed using the statistical software R (R Development Core Team, 2006) running under Linux (Red Hat 9.0).

Evaluation of the tests for genetic polymorphism effect. In the first step, each of the three tests presented is applied to detect the effect of the exon 26 polymorphism on the 1,000 data sets simulated under H_0 for each design (40 and 200 patients). Tests are performed on estimations obtained for each data set using FO and FOCE in NONMEM. The type I error for each analysis is defined as the percentage of data sets, where the corresponding test was significant. The expected prediction interval with 1,000 simulations and a value of 5% is [3.7; 6.3]. To ensure

a type I error of 5%, we define a correction threshold as the 5th percentile of the distribution of the p-values of the test under H_0.

In the second step, for the design with 40 patients, the same tests are performed using the 1,000 data sets simulated under H_1. The power is defined as the percentage of data sets, where the corresponding test was significant. We use the corrected threshold to compute the corrected powers. This allows comparison of the different tests even if the type I error is different from 5%.

Data sets with a group defined by the genotype for exon 26 with less than two patients are discarded from the analysis. It should be emphasized that the number of simulated data sets for which the tests can be applied are often less than 1,000. Indeed, the algorithms FO and FOCE used in the NONMEM software are sometimes unable to converge on a data set. In that case, estimates are not available and no tests can be applied. Similarly, even when the linearization algorithms achieve convergence, the variance covariance matrix is not always available. More precisely, to perform the ANOVA on the EBE, only the convergence of model M_{base} is required with or without the SE estimates. The Wald test requires the parameter estimation error to compute the test statistics, thus not only the convergence of model M_{mult} is required but the variance covariance matrix must also be available. The convergence of both M_{base} and M_{mult} is necessary to apply the LRT, but the covariance step is not needed to succeed. The three tests were also evaluated on a subset of data sets fulfilling all the conditions listed in this paragraph using FO or FOCE.

Evaluation of the strategies for model building. The different model building strategies described previously based either on tests or selection criteria are evaluated on data sets with the $N = 40$ design. Results are reported as the percentage of data sets for which each model of the exon 26 polymorphism effect is selected, and this with simulation under H_0 or under H_1. The correct model is M_{base} under H_0 and M_{mult} under H_1. The results of the strategies obtained by simulation under H_0 are not used to modify the strategies under H_1 because there are no simple correction methods for over-selecting covariate models under H_0.

As for the evaluation of tests, the analyses are not always performed on the 1,000 data sets. The conditions to perform model building following an ANOVA are identical to those described for the test. To choose the best model using Wald tests, the convergence and the covariance step are required for the three models including the gene effect. For the selection with the LRT, convergence of the four models is required, and for the criteria based selection, at least one model must have achieved convergence. The different strategies were also evaluated on samples satisfying all conditions with FO and FOCE.

Application to real data. The indinavir concentrations are analyzed with the same PK model as in the simulation study. The estimation method is chosen based on performances in the simulation study. Both the exons 26 and 21 of the ABCB1 gene are investigated using the three tests and the eight model building strategies.

RESULTS

Type I Error and Power of the Detection Tests

For the design with $N = 40$, three simulated data sets (one under H_0 and two under H_1) are discarded from the analysis due to a group of less than two

patients with the TT genotype. Also, FOCE encounters many more convergence problems than FO; K, the number of data sets among the 1,000 on which the test is performed, is always lower with FOCE.

Type I error estimates of the three tests performed are shown in Table 1. For the design with $N = 40$, the ANOVA type I errors do not significantly differ from 5% with both estimation algorithms. The LRT type I error estimates show a slight significant increase with FOCE and increase tenfold using FO. For the Wald test, there is a rather important significant increase for FOCE and again fourfold inflation with FO. For the design with 200 patients, the LRT attains a type I error nonsignificantly different from 5%, and the Wald test still has an estimate slightly superior to the nominal level with FOCE. The large increase of the type I error remains the same with FO for the LRT and the Wald test.

Estimates are given of the power of the tests using both estimation algorithms for the ANOVA, the Wald test, and the LRT for the design with 40 patients in Table 2. For each test and each estimation method, the corrected power is computed using the corresponding empirical threshold in order to maintain a type I error of 5%. The corrected power for LRT and Wald tests based on the FO method were low, but the inflated type I error in these situations already shows that this estimation method is poor. The powers for the ANOVA (using FO or FOCE) and the LRT for FOCE are around 70%, but the power of the Wald approach for FOCE is much lower (25%). We explored the FOCE outputs to understand this last result. In fact, with the FOCE algorithm in NONMEM, we observe correlations between the estimates of the gene effect coefficients and their estimation errors. This relationship leads to decreased values of the Wald statistic and therefore reduces the power to detect a genetic polymorphism effect. The same results were obtained considering only the subset of data sets fulfilling all convergence conditions for both FO and FOCE.

Model Selection Strategies

The results of the model selection strategies applied to the data sets simulated under H_0 with the design of 40 patients are presented in Table 3. Again, the performances of the FO algorithm are unsatisfactory except for the strategies based on EBE. With FOCE, both model selection strategies based on EBE and LRT select

Table 1 Type I error for each test and for each algorithm

		$N = 40$		$N = 200$	
Test	Algorithm	K	Type I error (%)	K	Type I error (%)
ANOVA	FO	997	5.9	1000	4.4
	FOCE	986	5.6	982	5.1
Wald	FO	975	23.4*	984	10.3*
	FOCE	924	11.7*	860	6.5*
LRT	FO	989	46.9*	977	54.0*
	FOCE	964	7.9*	956	5.0

K is the number of data sets on which the test could be performed.
*Estimate significantly different from 5%.

Table 2 Power for each test for $N = 40$

Test	Algorithm	K	Power (%)	Corrected power (%)
ANOVA	FO	995	69.8	66.5
	FOCE	968	71.2	69.3
Wald	FO	974	61.5	7.7
	FOCE	905	57.2	24.7
LRT	FO	958	90.2	48.7
	FOCE	947	78.7	71.0

K is the number of data sets on which the test could be performed.

The corrected power was obtained using the fifth percentile of the empirical distribution of the test statistic under H_0 as the cut-off value for the test.

the correct model M_{base} in about 90% of instances, whereas the Wald approach selects a model with a gene effect in around 15% of the data sets. Using selection criteria, the AIC and AICc obtain the worst performances, selecting a model with a gene effect in 57.8% and 51.7% of the data sets, respectively. The BICc shows performance close to that of the approach based on Wald tests, whereas CAIC (4.7%) and BIC (6.7%) select a model with a gene effect in about 5% of the data sets.

The results on data sets with $N = 40$ simulated under H_1 are given in Table 4. The results of the strategies using the LRT, Wald tests, or a criterion with the FO

Table 3 Percentage of data sets simulated under H_0 with a design of $N = 40$ for which each model is selected

			Model[1]			
Method	Algorithm	K	M_{base}	$M_{\text{recessive}}$	M_{dominant}	M_{mult}
ANOVA	FO	997	91.6	4.1	3.9	0.4
	FOCE	986	90.9	3.8	4.7	0.6
Wald	FO	947	68.4	11.1	16.3	4.2
	FOCE	876	83.0	5.8	9.8	1.4
LRT	FO	976	50.3	18.7	17.0	14.0
	FOCE	951	91.3	4.0	3.5	1.2
AIC	FO	999	14.9	23.2	22.4	39.5
	FOCE	970	42.2	22.4	21.3	14.1
AICc	FO	999	17.4	24.4	23.2	35.0
	FOCE	970	49.3	20.7	20.1	9.9
CAIC	FO	999	63.0	16.6	14.5	5.9
	FOCE	970	95.3	2.1	2.5	0.1
BIC	FO	999	55.7	19.6	16.6	8.1
	FOCE	970	93.3	3.1	3.0	0.6
BICc	FO	999	44.6	21.7	20.0	13.7
	FOCE	970	85.8	7.0	6.0	1.2

K is the number of data sets on which the test could be performed.

[1] $M_{\text{base}} : \{\beta_0 = \beta_1 = \beta_2 = 1\}$ (CC = CT = TT) model with no gene effect.

$M_{\text{recessive}} : \{\beta_0 = \beta_1 = 1, \beta_2 \neq 1\}$ (CC = CT ≠ TT), reduced model.

$M_{\text{dominant}} : \{\beta_0 = 1, \beta_1 = \beta_2 \neq 1\}$ (CC ≠ CT = TT), reduced model.

$M_{\text{mult}} : \{\beta_0 = 1, \beta_1 \neq \beta_2 \neq 1\}$ (CC ≠ CT ≠ TT), complete model.

Table 4 Percentage of data sets simulated under H_1 with $N = 40$ for which each model is selected

Method	Algorithm	K	Model[1]			
			M_{base}	$M_{recessive}$	$M_{dominant}$	M_{mult}
ANOVA	FO	995	25.5	44.8	21.0	8.7
	FOCE	968	22.3	43.3	25.5	8.9
Wald	FOCE	878	17.9	41.7	32.7	7.7
LRT	FOCE	923	19.1	47.3	20.5	13.1
AIC	FOCE	962	1.3	31.1	13.1	54.5
AICc	FOCE	962	1.6	35.1	14.7	48.6
CAIC	FOCE	962	28.1	48.5	19.9	3.5
BIC	FOCE	962	21.6	50.3	21.7	6.4
BICc	FOCE	962	11.5	51.7	23.1	13.7

K is the number of data sets on which the test could be performed.
[1]$M_{base} : \{\beta_0 = \beta_1 = \beta_2 = 1\}$ (CC = CT = TT) model with no gene effect.
$M_{recessive} : \{\beta_0 = \beta_1 = 1, \beta_2 \neq 1\}$ (CC = CT ≠ TT), reduced model.
$M_{dominant} : \{\beta_0 = 1, \beta_1 = \beta_2 \neq 1\}$ (CC ≠ CT = TT), reduced model.
$M_{mult} : \{\beta_0 = 1, \beta_1 \neq \beta_2 \neq 1\}$ (CC ≠ CT ≠ TT), complete model.
Results obtained with FO are not presented for these strategies because of their poor performance under H_0 (Table 3).

estimation method are not presented because of the poor properties under H_0. A model with a gene effect is selected for about 70% of the data sets using CAIC, for about 80% of the data sets using the different model-building strategies based on tests or BIC, for about 90% of the data sets using BICc, and for about 99% of the data sets using AIC and AICc. However, for the latter, the percentage of data sets where M_{base} is not selected under H_0 is greater than 50%. Another noticeable result is that the model used to simulate the data M_{mult} is seldom selected compared with the intermediate model $M_{recessive}$, which is chosen in 30 to 50% of the data sets using the different methods. The simulated value β_1 is low compared with β_2, therefore if the model-building strategies succeed in selecting a model with a gene effect, it is not always the correct one. The performances are similar using the sample of data sets for which all conditions were put together with FO and FOCE.

Figure 2 represents a summary of these results, placing each strategy on a bi-dimensional plan; the ability to select a model with a gene effect under H_1 vs. the trend not to select M_{base} under H_0. The CAIC is the test that has the best properties under H_0, but the BIC more often selects a model with covariate under H_1, whereas it is only slightly less conservative under H_0. We also note that the LRT, the BIC, and the ANOVA are clustered together and thus offer similar compromises.

Application

Figure 3 represents a spaghetti plot of concentrations vs. time for the 37 patients from the indinavir arm of the COPHAR2 study, sorted by genotype classes for exon 26. Concentrations show an important interindividual variability and only three patients were mutant homozygotes for this polymorphism.

As the simulation study has shown poor performances with the FO algorithm, we estimated the model parameters using the FOCE with interaction algorithm.

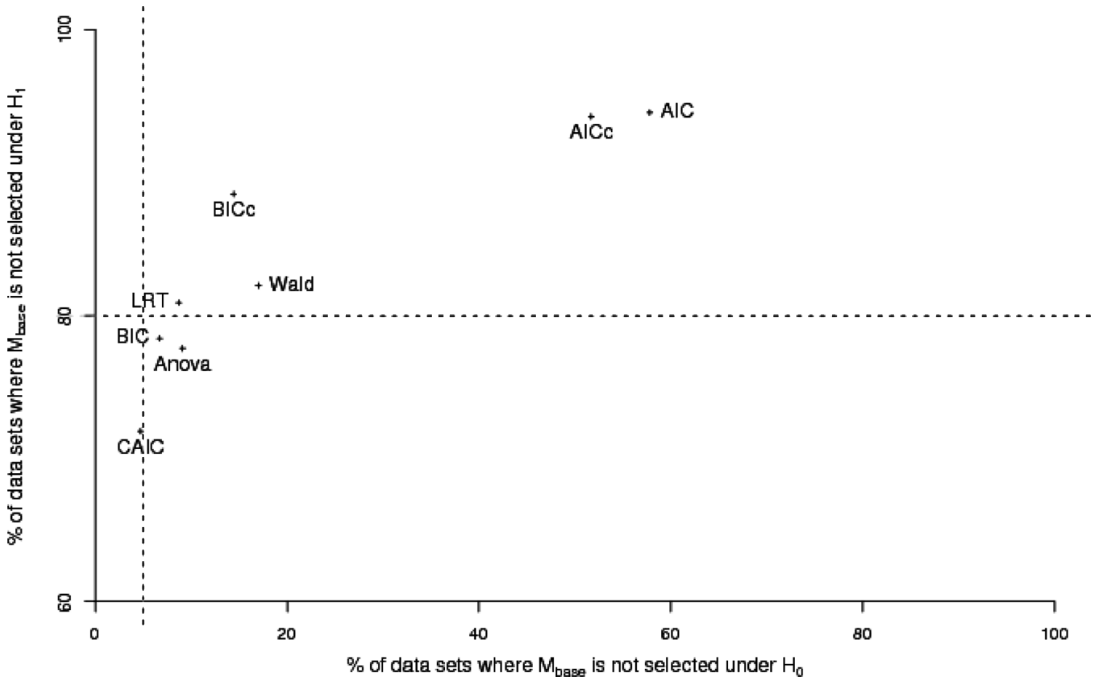


Figure 2 Percentage of data sets simulated under H_1 where the model with no gene effect (M_{base}) is not selected vs. the same percentage under H_0 for the eight model selection strategies using FOCE for $N = 40$. The vertical line corresponds to a value of 5% and the horizontal line to a value of 80%.

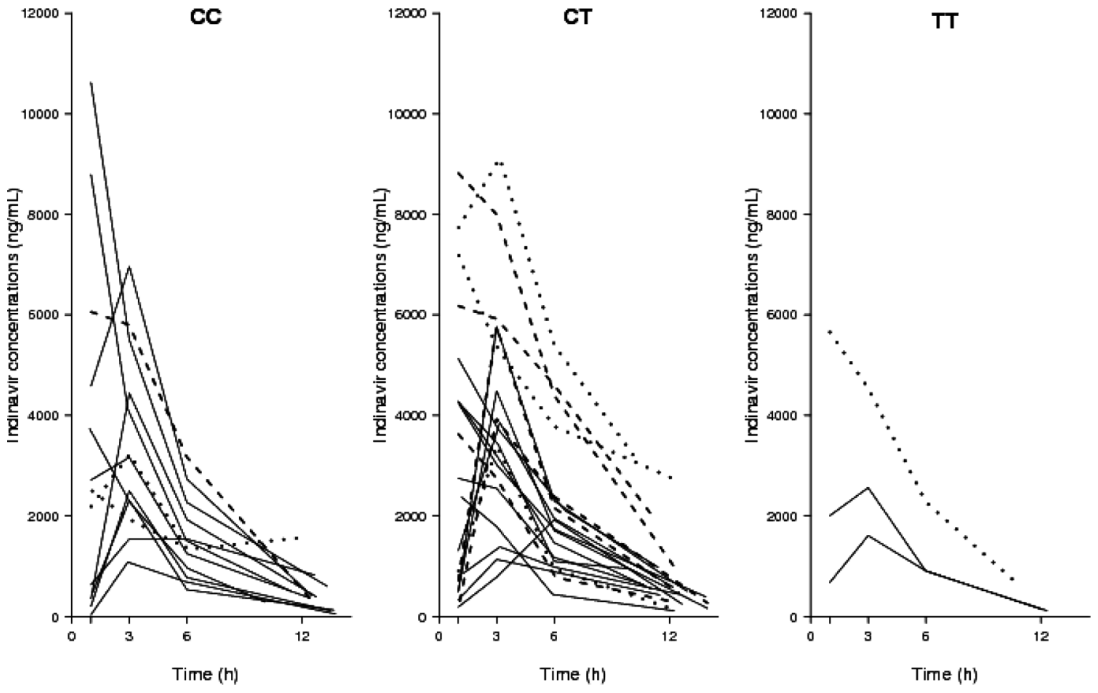


Figure 3 Indinavir concentrations (ng/mL) at steady state collected in the COPHAR2-ANRS 111 trial vs. time, sorted by ABCB1 exon 26 genotypes. The plain lines correspond to a dose of 400 mg indinavir, the dashed lines correspond to 600 mg, and the dotted lines correspond to 800 mg, all with a dose of 100 mg ritonavir bid.

Therefore, the estimates are different from the simulated values obtained in the preliminary analysis. In addition, the estimate of the interindividual variability for k was very small and we fixed it to 0 (no variability). The model with no covariate had an absorption constant of $0.8\,h^{-1}$ with an important interindividual variability of 70.3%, an elimination constant of $0.2\,h^{-1}$, and a volume of distribution of 99.3 L with an interindividual variability of 47%. All the estimation errors were below 20% for the fixed effects and below 40% for the variances.

For the influence of the ABCB1 exon 26 on the indinavir volume of distribution, the ANOVA and the LRT were nonsignificant ($P = 0.7$ and 0.2, respectively). The global Wald test obtained a p-value of 0.02, however the corrected threshold defined in the simulation study for this test is $7.5 \cdot 10^{-3}$. The model with no covariate was chosen using the selection strategies based on the EBE, or on the LRT, using CAIC, BIC, and BICc, whereas the strategy based on Wald tests, AIC, and AICc selected $M_{\text{recessive}}$. No influence of the ABCB1 exon 21 on indinavir volume of distribution was detected using any of the tests or selection strategies under study.

DISCUSSION

In this work, we evaluate several statistical tests and model selection strategies using nonlinear mixed effects models to analyze the impact of a genetic polymorphism on one PK parameter through simulation. We also study the impact of the estimation algorithms in NONMEM, comparing the two first-order approximations most widely used, FO and FOCE, which linearize the model function around the random effects equal to 0 and around the individual estimates of the random effects, respectively. Although the FO method has been shown to suffer from various problems, it is still used because the FOCE algorithm is known for numerical difficulties and for its slowness. Using the FO algorithm we observe unsatisfactory performances for all the tests and model selection strategies with the exception of methods based on the EBE. The linearization of the likelihood function around the fixed effects leads to type I error inflation (Comets and Mentré, 2001; Panhard and Mentré, 2005; Wählby et al., 2001).

With FOCE, there is a significant increase in the type I error of the LRT and the Wald approach with a design including 40 patients. This increase has already been described for the LRT (Comets and Mentré, 2001; Wählby et al., 2001) and has also been shown for the Wald test (Panhard and Mentré, 2005). The design with 200 patients is closer to asymptotic conditions and shows as expected a correct type I error. Performing simulations under H_0 can be used to correct the threshold for the test under H_1 as we did in the present study. We observe a power around 70% for the tests using ANOVA and LRT, which is close to the power of 80% expected from the simulation settings. Of course if the study was designed specifically to detect a gene effect the sample size could be increased to ensure a higher (more ethical) power. Our objective here was to compare different methods, and we used the data from the COPHAR2-ANRS 111 trial to provide the settings for the PK simulations, as we have analyzed these data thereafter. The simulated effect of gene was chosen to be consistent with the literature, though in the COPHAR2 trial the genotype distribution is slightly unbalanced. The reduced power of the Wald approach and the unsatisfactory efficacy of the model selection strategy based on the Wald tests could result partly from a wrong estimation of the standard errors

due to the log-likelihood function linearization. Indeed, we observe with FOCE that in the simulations the estimation errors are highly correlated to their estimates. The Wald statistic is based on the ratio of the estimates to the estimation error, and this could explain the poor performance of the Wald tests under H_1. Finally, FOCE met with convergence problems: on M_0 and $M_{\text{CCvsCTvsTT}}$, which involved no mathematical complexities for $N = 40$ under H_1, 35 (3.5%) and 48 (4.8%) runs did not achieve convergence, either for numerical reasons or because they had to be terminated and we could not obtain an estimate of the variance–covariance matrix in 32 (3.2%) and 38 (3.8%) runs, respectively. This could have been improved partly by the use of different initial conditions, nonetheless the results were identical using the sample of data sets fulfilling convergence conditions for the three tests using both estimation methods. Among the different implementations of both algorithms available, we chose to use NONMEM because it is the most popular tool in the pharmaceutical industry. Our results can be extended to the FO method implemented in SAS because it computes the same likelihood function up to a constant (Wang, 2007), although it is not the same algorithm (Roe, 1997). Similarly, FOCE with interaction and the NLME method with the varConstPower option implemented in Splus should give the same results, save for the lower accuracy of the Splus approximation (Girard and Mentré, 2005; Wang, 2007). The simulations and analyses in the present paper have been performed using version 5.1 of NONMEM. A new version of NONMEM, version 6, was released in December 2006 and was implemented in our department after the major part of the analyses had already been run, so that we kept NONMEM 5.1 for this study. The FOCE routine in NONMEM 6 has been rewritten and should provide more stable runs, which may reduce the convergence issues we have found in the present study. However, in our preliminary results using NONMEM 6 on a subset of the simulated datasets, an increased number of runs failed the covariance-step. Rather than reporting partial results in the present study, we will investigate this matter in a subsequent work.

Model selection strategies based on tests have, to a certain extent, a high rate of false inclusion under H_0 (over 10%), which could result from the uncorrected multiple model comparisons. Further, there is no simple way to correct the model selection under H_1 by taking into account the simulations under H_0. Consequently, we have to be cautious about assessing the performance under H_1 when the behavior is poor under H_0.

With respect to selection based on criteria, except for the performances of the BIC and the BICc not showing any trend to conservatism, under H_0 our results agree with the literature. The acceptable performance of the CAIC has to be noted, as well as the very poor performance of AIC and AICc. Under H_1, there is a satisfactory weak selection of M_{base} with a rather important representation of $M_{\text{recessive}}$. Indeed, in our simulation conditions, $M_{\text{recessive}}$ and M_{mult} are close, and the power to detect a difference between these two models is much lower than the power to detect a difference between M_{base} and M_{mult}. As a side-note, in the Bayesian literature, it is usual also to consider not only the best model (i.e., the model with the lowest criterion) but also models close to the best model (Raftery, 1995). However, in our study the simulation model is rarely close in this sense to the best model.

Finally, in our study, for the design with 40 patients, the ANOVA on the EBE is the only test that maintained a 5% type I error as well as a good power. It should be noted that we simulated a sufficient number of samples per patient with respect

to the number of model parameters. In sparse sample situations, regression to the mean is known to occur with EBE (Panhard and Mentré, 2005), which could result in lower power to detect differences between genotype groups. We plan to test this hypothesis in a subsequent study with a sparse design including only two samples per patient.

We also confirm that both AIC and AICc should not be used for model building but further studies are required to provide recommendations on the other selection strategies. Regarding the estimation methods, if FO can still be used in covariate screening on the EBE, one should avoid performing model building with the LRT or Wald tests based on results from this algorithm.

We illustrate the different approaches using data from the indinavir PK sub-study of the COPHAR2 ANRS-111 trial. The PK model has already been described but no gene effect has been investigated using a population approach to date. The estimated parameters are in accordance with estimations obtained in other studies (Csajka et al., 2004; Goujard et al., 2005; Kappelhoff et al., 2005). The Wald test is the only test to detect an influence of ABCB1 exon 26 on the volume of distribution. Considering the corrected threshold provided by the simulation study, we can probably ascribe this discrepant result to the inflated type I error. Similarly, it is the three strategies with the highest percentage of data sets where M_{base} is not selected under H_0 that select a model with the ABCB1 exon 26 polymorphism as covariate. The polymorphism on the exon 26 of the ABCB1 gene has been shown to impact on plasma concentrations of nelfinavir, another protease inhibitor (Fellay et al., 2002). However, this work agrees with another study (Verstuyft et al., 2005), where no effect of the polymorphisms from ABCB1 exon 26 and 21 on indinavir bioavailability was found.

Another extension of this work would be to simulate under H_1 various levels of the gene effects, which would provide more information on the relationship between the strength of the genetic polymorphism effect and the power. Designing an optimal sampling schedule for testing a gene effect with a given power is also an interesting challenge, and the extension of the PFIM software for design optimization in the case of models with covariates could be used (Retout and Mentre, 2003; Retout et al., 2007). Moreover in the genetic framework one should keep in mind the complex pathway leading from DNA to metabolic activity, which is usually controlled by more than one exon. Another perspective would therefore be to analyze the influence of the haplotypes (Inomata et al., 2005), because such a classification seems to be more relevant at the DNA level.

Pharmacogenetic studies using NLMEM have many advantages because fewer samples are required to estimate parameters with a biological meaning. The current literature presents a wide array of methods for covariate selection using NLMEM. We show in this study that methods using EBE are efficient not only in data exploration but also in model selection on data sets with enough samples per patient. We also emphasize that using estimation algorithms based on likelihood linearization, LRT type I error is inflated, thus one has to perform simulations or work with large data sets. Finally, the problems of the FO and FOCE algorithm in terms of convergence and bias are an incentive to use more recent estimation methods (Samson et al., 2007).

ACKNOWLEDGMENTS

We would like to thank the COPHAR 2-ANRS 111 scientific committee (investigators Pr. D. Salmon and Dr. X. Duval, pharmacology; Pr. J. M. Tréluyer, methodology; and Pr. F. Mentré) for giving us access to the PK data of the indinavir arm in order to build our simulations and to illustrate our topic. We would also like to thank the IFR02 of INSERM and Hervé Le Nagard for the use of the "centre de biomodélisation."

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19:716–723.

Beal, S. L. (2005). Conditioning on certain random events associated with statistical variability in PK/PD. *J. Pharmacokinet. Pharmacodyn.* 32:213–243.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Met.* 57:289–300.

Bodor, M., Kelly, E. J., Ho, R. J. (2005). Characterization of the human MDR1 gene. *AAPS J.* 7:E1–E5.

Bozdogan, H. (1987). Model selection and Akaike's information criteria (AIC): The general theory and its analytical extensions. *Psychometrika* 52:345–370.

Brendel, K., Legrand, M., Taburet, A. M., Baron, G., Goujard, C., Mentre, F. (2005). Population pharmacokinetic analysis of indinavir in HIV-infected patient treated with a stable antiretroviral therapy. *Fundam. Clin. Pharmacol.* 19:373–383.

Burnham, K. P., Anderson, D. R. (2002). *Model Selection and Multimodel. A Practical Information – Theoric Approach*. New York: Springer-Verlag.

Comets, E., Mentré, F. (2001). Evaluation of tests based on individual vs population modelling to compare dissolution curves. *J. Biopharm. Stat.* 11:107–123.

Comets, E., Verstuyft, C., Lavielle, M., Jaillon, P., Becquemont, L., Mentré, F. (2007). Modelling the influence of MDR1 polymorphism on digoxin pharmacokinetic parameters. *Eur. J. Pharmacol.* 63:437–449.

Csajka, C., Marzolini, C., Fattinger, K., Decosterd, L. A., Telenti, A., Biollaz, J., Buclin, T. (2004). Population pharmacokinetics of indinavir in patients infected with human immunodeficiency virus. *Antimicrob. Agents Chemother.* 48:3226–3232.

Fellay, J., Marzolini, C., Meaden, E., Back, D., Buclin, T., Chave, J. (2002). Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter gene MDR1: A pharmacogenetic study. *Lancet* 359:30–36.

Gabrielsson, J., Weiner, D. (1999). *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*. Stockholm: Apotekarsocieteten.

Girard, P., Mentré, F. (2005). A comparison of estimation methods in nonlinear mixed effects models using a blind analysis. *PAGE* 14:Abstr 834.

Goujard, C., Legrand, M., Panhard, X., Diquet, B., Duval, X., Peytavin, G., Vincent, I., Katlama, C., Leport, C., Bonnet, B., Salmon-Ceron, D., Mentre, F., Taburet, A. M. (2005). High variability of indinavir and nelfinavir pharmacokinetics in HIV-infected patients with a sustained virological response on highly active antiretroviral therapy. *Clin. Pharmacokinet.* 44:1267–1278.

Henningsson, A., Marsh, S., Loos, W. J., Karlsson, M. O., Garsa, A., Mross, K., Mielke, S., Vigano, L., Locatelli, A., Verweij, J., Sparreboom, A., McLeod, H. L. (2005). Association of CYP2C8, CYP3A4, CYP3A5, and ABCB1 polymorphisms with the pharmacokinetics of paclitaxel. *Clin. Cancer Res.* 11:8097–8104.

Innocentia, F., Liua, W., Chenb, P., Desaia, A. A., Dasb, S., Rataina, M. J. (2005). Haplotypes of variants in the UDP-glucuronosyltransferase 1A9 and 1A1 genes. *Pharmacogenet. Genomics* 15:295–301.

Inomata, S., Nagashima, A., Itagaki, F., Homma, M., Nishimura, M., Osaka, Y., Okuyama, K., Tanaka, E., Nakamura, T., Kohda, Y., Naito, S., Miyabe, M., Toyooka, H. (2005). CYP2C19 genotype affects diazepam pharmacokinetics and emergence from general anesthesia. *Clin. Pharmacol. Ther.* 78:647–655.

Kappelhoff, B. S., Huitema, A. D., Sankatsing, S. U., Meenhorst, P. L., Van Gorp, E. C., Mulder, J. W., Prins, J. M., Beijnen, J. H. (2005). Population pharmacokinetics of indinavir alone and in combination with ritonavir in HIV-1-infected patients. *Br. J. Clin. Pharmacol.* 60:276–286.

Kerbusch, T., Wählby, U., Milligan, P., Karlsson, M. (2003). Population pharmacokinetic modelling of darifenacin and its hydroxylated metabolite using pooled data, incorporating saturable first pass metabolism, CYP2D6 genotype and formulation-dependant bio availability. *Br. J. Clin. Pharmacol.* 56:639–652.

Licinio, J., Wong, M. L. (2002). *Pharmacogenomics: The Search for Individualized Therapy*. Weinheim: Wiley-VCH.

Machin, D., Campbell, M., Fayers, P., Pinol, A. (1997). *Sample Size Tables for Clinical Studies*. Oxford: Blackwell Science.

Mamiya, K., Hadama, A., Yukawa, E., Ieiri, I., Otsubo, K., Ninomiya, H., Tashiro, N., Higuchi, S. (2000). CYP2C9 polymorphism effect on phenobarbiton kinetics in Japanese patients with epilepsy: Analysis by population phamacokinetics. *Eur. J. Clin. Pharmacol.* 55:821–825.

Marzolini, C., Paus, E., Buclin, T., Kim, R. B. (2003). Polymorphisms in Human MDR1 (P-glycoprotein): Recent advances and clinical relevance. *Clin. Pharmacol. Ther.* 75:13–33.

Meerum, T. J., Malingre, M., Beijnen, J., Ten, B., Huinink, W., Rosing, H., Koopman, F., et al. (1999). Co-administration of cyclosporine A enables oral therapy with paclitaxel. *Clin. Cancer Res.* 5:3379–3384.

Mentré, F., Duval, X., Rey, E., Peytavin, G., Auleley, S., Legrand, M., Biour, M., Bouxin, A., Goujard, C., Taburet, A., Katlama, C., Diquet, B., Lascoux, C., Tréluyer, J. M., Salmon-Céron, D. (2005). Prospective Trial to Evaluate How Therapeutic Drug Monitoring of Protease Inhibitors Increases Virologic Success and Tolerance of HAART (COPHAR 2-ANRS 111 Trial). 12th Conference on Retroviruses and Opportunistic Infections. Boston.

Min, D. I., Ellingrod, V. L., Marsh, S., McLeod, H. (2004). CYP3A5 polymorphism and the ethnic differences in cyclosporine pharmacokinetics in healthy subjects. *Ther. Drug Monit.* 26:524–528.

Panhard, X., Mentré, F. (2005). Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence cross-over trials. *Stat. Med.* 24:1509–1524.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). *Sociol. Methodol.* 111–195.

Retout, S., Comets, E., Samson, A., Mentre, F. (2007). Design in nonlinear mixed effects models: Optimization using the Fedorov–Wynn algorithm and power of the Wald test for binary covariates. *Stat. Med.* 26:5162–5179.

Retout, S., Mentre, F. (2003). Optimization of individual and population designs using Splus. *J. Pharmacokinet. Pharmacodyn.* 30:417–443.

Roe, D. J. (1997). Comparison of population pharmacokinetic modeling methods using simulated data: results from the Population Modeling Workgroup. *Stat. Med.* 16:1241–1257; discussion 1257–1262.

Sakaeda, T., Nakamura, T., Okumura, K. (2002). MDR1 genotype-related pharmacokinetics and pharmacodynamics. *Biol. Pharm. Bull.* 25:1391–1400.

Samson, A., Lavielle, M., Mentré, F. (2007). The SAEM algorithm for group comparison test in longitudinal data analysis based in nonlinear mixed-effects model. *Stat. Med.* 26:4860–4875.

Schwartz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6:461–464.

Sheiner, L., Beal, S. (1998). *NONMEM Version 5.1*. San Francisco: University of California, NONMEM Project Group.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communication in Statistics. Theory and Methods* 7:13–26.

Taguchi, M., Fujiki, A., Iwamoto, J., Inoue, H., Tahara, K., Saigusa, K., Horiuchi, I., Oshima, Y., Hashimoto, Y. (2006). Nonlinear mixed effects model analysis of the pharmacokinetics of routinely administered bepridil in japanese patients with arrhythmias. *Biol. Pharm. Bull.* 29:517–521.

Verstuyft, C., Marcellin, R., Morand-Joubert, L., Launay, O., Brendel, K., Mentré, F., Peytavin, G., Gérard, L., Becquemont, L., Aboulker, J. P. (2005). Absence of association between MDR1 genetic polymorphisms, indinavir pharmacokinetics and response to HAART. *AIDS* 19:2127–2131.

Wählby, U., Jonsson, E. N., Karlsson, M. O. (2001). Assessment of actual significance levels for covariate effects in NONMEM. *J. Pharmacokinet. Pharmacodyn.* 28:231–252.

Wang, Y. (2007). Derivation of various NONMEM estimation methods. *J. Pharmacokinet. Pharmacodyn.* 34:575–593.

Yamaguchi, H., Hishinuma, T., Endo, N., Tsukamoto, H., Kishikawa, Y., Sato, M., Murai, Y., Hiratsuka, M., Ito, K., Okamura, C., Yaegashi, N., Suzuki, N., Tomioka, Y., Goto, J. (2006). Genetic variation in ABCB1 influences paclitaxel pharmacokinetics in Japanese patients with ovarian cancer. *Int. J. Gynecol. Cancer* 16:979–985.

2.2 Impact du plan d'expérience sur les tests de détection d'un effet gène sur un paramètre pharmacocinétique

2.2.1 Résumé

Dans l'étude précédente nous avons montré que seule l'ANOVA conservait une erreur de type I non significativement différente de 5% sur le plan d'expérience inspiré de l'étude réelle. En effet, le test de Wald et le LRT montraient une inflation importante avec FO et plus légère mais significative avec FOCE lorsque le nombre de sujets est réduit. Les deux méthodes reposant sur une linéarisation du modèle, rien ne garantit la consistance des estimateurs ainsi obtenus (Vonesh, 1996).

Nous avons donc voulu déterminer si la méthode d'estimation exacte SAEM obtiendrait des résultats différents des algorithmes basés sur une linéarisation dans ce contexte. De plus, les recommandations émises par les autorités de santé (Committee for human medicinal products, 2007) sur la nécessité d'augmenter le nombre de sujets nous ont amené à explorer deux plans d'expérience alternatifs permettant d'augmenter le nombre de sujets dans chaque génotype. Le premier plan d'expérience inclut N=80 patients répartis en 4 groupes avec n=2 prélèvements par patient à des temps différents déterminés parmi les quatre temps de l'étude réelle. Pour optimiser la répartition des temps et les proportions des groupes, nous avons utilisé le logiciel PFIM Interface version 2.1 (Retout et al., 2007a). Ce logiciel utilise un algorithme de Fedorov-Wynn qui maximise le déterminant de la matrice de Fisher sur un ensemble fini de plans d'expérience, pour un modèle et des paramètres de population pré-définis. Le second plan d'expérience est une alternative réalisable partiellement en routine clinique, incluant N=20 sujets avec la cinétique complète de l'étude réelle et N=80 sujets ayant seulement une concentration résiduelle. Nous avons aussi réalisé une revue de la littérature sur les études de simulation évaluant les tests en pharmacocinétique/pharmacodynamie de population.

Sur le premier plan d'expérience, l'algorithme SAEM obtient des résultats équivalents à l'algorithme FOCE en termes d'erreur de type I et de puissance des tests, à l'exception de la puissance du test de Wald. Cependant, nous avons pu obtenir des estimations sur l'ensemble des jeux de données simulés et nous avons rencontré bien moins de problèmes numériques qu'avec FOCE. L'ANOVA conserve une erreur de type I non significativement différente sur les quatre plans d'expérience à l'étude, alors que le test de Wald et le LRT rencontrent une inflation significative sur les 3 plans d'expérience avec un total de cent soixante observations. Cette inflation est associée à une sous-estimation des erreurs d'estimation des coefficients d'effet, et donc à un éloignement aux conditions asymptotiques requises pour ces tests. En effet, avec les erreurs d'estimation empiriques (plus élevées), l'erreur de type I du test est non significativement différente

du seuil nominal. Pourtant, les études de simulation retrouvées dans la littérature et ayant des plans d'expériences comparables (N=40/n=4) n'ont pas observé d'inflation de l'erreur de type I de ces tests. Il semble donc que ce soit le nombre de classes et la distribution déséquilibrée de la covariable génétique (1/4, 1/2 et 1/4 respectivement pour les homozygotes communs, les hétérozygotes et les homozygotes rares) qui entraînent cette inflation.

Les MNLEM permettent d'évaluer des plans d'expériences plus flexibles, avec moins de prélèvements par sujet, et ainsi d'augmenter le nombre de sujets dans chaque groupe de génotype. En conséquence, une grande diversité de choix de protocoles est observée dans les études pharmacogénétiques, comme nous l'avons décrit dans la partie 1.5.4. Cependant, lorsque le nombre de sujets et/ou de prélèvements par sujet est limité et que certains génotypes sont sous-représentés, il faut réaliser une ANOVA ou utiliser une correction pour l'inflation de l'erreur de type I des tests asymptotiques.

L'ensemble de ce travail est décrit dans un article publié dans la revue *Journal of Pharmacokinetics and Pharmacodynamics*.

2.2.2 Article 2 (publié)

J Pharmacokinet Pharmacodyn (2009) 36:317–339
DOI 10.1007/s10928-009-9124-x

Pharmacogenetics and population pharmacokinetics: impact of the design on three tests using the SAEM algorithm

Julie Bertrand · Emmanuelle Comets · Céline M. Laffont · Marylore Chenel · France Mentré

Received: 21 February 2009 / Accepted: 17 June 2009 / Published online: 27 June 2009

Abstract Pharmacogenetics is now widely investigated and health institutions acknowledge its place in clinical pharmacokinetics. Our objective is to assess through a simulation study, the impact of design on the statistical performances of three different tests used for analysis of pharmacogenetic information with nonlinear mixed effects models: (i) an ANOVA to test the relationship between the empirical Bayes estimates of the model parameter of interest and the genetic covariate, (ii) a global Wald test to assess whether estimates for the gene effect are significant, and (iii) a likelihood ratio test (LRT) between the model with and without the genetic covariate. We use the stochastic EM algorithm (SAEM) implemented in MONOLIX 2.1 software. The simulation setting is inspired from a real pharmacokinetic study. We investigate four designs with N the number of subjects and n the number of samples per subject: (i) N = 40/n = 4, similar to the original study, (ii) N = 80/n = 2 sorted in 4 groups, a design optimized using the PFIM software, (iii) a combined design, N = 20/n = 4 plus N = 80 with only a trough concentration and (iv) N = 200/n = 4, to approach asymptotic conditions. We find that the ANOVA has a correct type I error estimate regardless of design, however the sparser design was optimized. The type I error of the Wald test and LRT are moderatly inflated in the designs far from the asymptotic (<10%). For each design, the corrected power is analogous for the three tests. Among the three designs with a total of 160 observations, the design N = 80/n = 2 optimized with PFIM provides both the lowest standard error on the effect coefficients and the best power

J. Bertrand (✉) · E. Comets · F. Mentré
UMR 738, INSERM, Université Paris Diderot, 16 rue Henri Huchard, 75018 Paris, France
e-mail: julie.bertrand@inserm.fr

C. M. Laffont
UMR181, Physiologie et Toxicologie Expérimentales INRA, ENVT, Toulouse, France

M. Chenel
Institut de Recherches Internationales Servier, Courbevoie, France

for the Wald test and the LRT while a high shrinkage decreases the power of the ANOVA. In conclusion, a correction method should be used for model-based tests in pharmacogenetic studies with reduced sample size and/or sparse sampling and, for the same amount of samples, some designs have better power than others.

Keywords Pharmacogenetics · Pharmacokinetics · Nonlinear mixed effects models · Test · Design · Single nucleotid polymorphism · SAEM

Introduction

Pharmacogenetics (PG) studies the influence of variations in DNA sequence on drug absorption, disposition and effects [1, 2]. This area is now widely investigated and the European Medicines Agency (EMEA) has published in 2007 a reflection paper acknowledging the place of PG in clinical pharmacokinetics (PK) [3].

Pharmacogenetic data are mainly studied using non-compartmental methods followed by a one-way analysis of variance (ANOVA) on the individual parameters of interest [4]. More sophisticated approaches have also been used such as NonLinear Mixed Effects Models (NLMEM). These models allow to integrate the knowledge accumulated on the drug PK, and they have the advantage of being applicable with less samples per patient.

Various methods can be used to include pharmacogenetic information in NLMEM. Preliminary screening is usually performed using ANOVA on the individual parameters estimates [5] followed by a stepwise model building approach with the likelihood ratio test (LRT) [6]. As an alternative approach, a global Wald test can assess whether estimates for the genetic effect are significant [7].

In a previous work [8], we performed a simulation study to assess the statistical properties of these different approaches. We used the estimation algorithms FO and FOCE interaction (FOCE-I) implemented in the NONMEM software version V [9]. In the present work, to avoid the linearisation step we use the Stochastic EM algorithm (SAEM), implemented in the MONOLIX software version 2.1 [10] for the analysis of the simulated data sets with the same three tests. SAEM computes exact maximum likelihood estimates of the model parameters using a stochastic version of the EM algorithm including a MCMC procedure.

In [8], we have simulated a design of 40 subjects inspired from a real pharmacokinetic substudy on indinavir performed during the COPHAR2-ANRS 111 trial in HIV patients [11, 12]. We have also simulated the same sampling schedule but with a larger sample size of 200 subjects to be closer to the asymptotic properties of the test. Whereas the estimated type I error of the ANOVA was found to be close to 5% whatever the design, those of the Wald test and the LRT showed for the FOCE-I algorithm a slight and significant increase, respectively, for the first design with 40 subjects. In the present paper, we aim to further investigate the impact of the design on the performances of these three tests in terms of type I error and power. The EMEA has stated that pharmacogenetic studies should include a satisfactory number of patients of each geno- or phenotype in order to obtain valid

correlation data [3]. Therefore, with the SAEM algorithm, we also consider two other designs with a larger number of subjects but different blood sampling strategies, as extensive sampling on each patient would no longer be practical. One of these designs was optimized using the PFIM interface software version 2.1 [13, 14] and another includes a group with only trough concentrations to explore a design that is easily implemented in practice. These two designs involve the same total number of observations as the original design with 40 subjects, to allow proper comparisons between designs.

In the first section of the article, we introduce the model as well as the notations, the three tests under study and the four designs. Then, we describe the simulation study and how we perform the evaluation. Next the main results of the simulation are exposed. Finally the study results and perspectives are discussed.

Methods

Model and notations

In this work, we consider the effect on a pharmacokinetic parameter of one biallelic Single Nucleotid Polymorphism (SNP), i.e. the existence of 2 variants for a base at a given locus on the gene. We denote, without loss of generality, C the wild allele and T the mutant, leading to k = 3 possible genotypes (CC, CT and TT). Let $y_{i,j}$ represents the concentration at time $t_{i,j}$ of a subject i = 1,…,N with genotype G_i at measurement j = 1,…,n such as:

$$y_{i,j} = f\left(t_{i,j}, G_i, \theta_i\right) + \epsilon_{i,j} \tag{1}$$

with θ_i the subject specific parameters of the nonlinear model function f and $\varepsilon_{i,j}$ the residual error normally distributed with zero mean and an heteroscedastic variance $\sigma^2_{i,j}$, with:

$$\sigma^2_{i,j} = \sigma^2(a + bf(t_{i,j}, G_i, \theta_i))^c \tag{2}$$

This combined error model (additive and proportional) is commonly used in population pharmacokinetics with c fixed to 2. For identifiability purpose σ^2 is set to one. We assume that the genetic polymorphism G_i for subject i affects θ_p, the pth component of the vector θ through the following relationship:

$$\theta_{p,i} = \mu_p e^{\beta_{G_i}} e^{\eta_{p,i}} \tag{3}$$

where μ_p is the population mean for parameter θ_p and $\eta_{p,i}$ follows a Gaussian distribution with zero mean and variance ω^2_p the pth diagonal element of matrix Ω. β_{G_i} is the effect coefficient corresponding to the genotype of subject i, we assume $\beta_{G_i} = 0, \beta_1$ or β_2 for G_i = CC, CT or TT, taking CC as the reference group.

In the following, we note M_{base} the model without a gene effect, where $\{\beta_1 = \beta_2 = 0\}$ i.e. $\{CC = CT = TT\}$, and M_{mult} the model with a multiplicative effect on the population mean of the parameter of interest, where $\{\beta_1 \neq \beta_2 \neq 0\}$ i.e. $\{CC \neq CT \neq TT\}$.

As in NLMEM the integral in the likelihood has no analytical form, specific algorithms are needed to estimate the model parameters and their standard error (SE) [15]. Since the beginning of the 21st century, EM-like algorithms appear as a potent alternative to the linearisation used in the earlier approaches. The SAEM algorithm is a stochastic version of EM algorithm where the individual parameter estimates are considered as the missing values [16]. The estimation step is decomposed in the simulation of the individual parameters using a Monte Carlo Markov Chain (MCMC) approach followed by the computation of stochastic approximation for some sufficient statistics of the model. The subsequent maximisation step of the sufficient statistics provides an update of the estimates. The estimation variance matrix is deduced from the NLMEM after linearisation of the function f around the conditional expectation of the individual parameters, the gradient of f being numerically computed.

The loglikelihood is obtained through importance sampling once parameter estimation is achieved, as follows. For each subject, $s = 1,\ldots,T$ samples of individual parameters are generated from a Gaussian approximation of the subject's individual posterior distribution. These T samples are used to derive T realizations of the loglikelihood, each weighted by the probability of the corresponding sample. The importance sampling estimator is the empirical average over the weighted T realizations. The variability of this approximation decreases when increasing the number of samples T [17].

Tests

Analysis of variance (ANOVA)

The data are analysed with the model not including the gene effect, M_{base}. We used the conditional expectation (mean) of the individual parameters provided by the MCMC procedure in SAEM as the empirical Bayes estimates (EBE). Then, the equality of the mean between the three genotypes is tested with an analysis of variance. The statistic is compared to the critical value of a Fisher distribution (F-distribution) with $3 - 1 = 2$ numerator degrees of freedom and $N - 3$ denominator degrees of freedom, 3 being the number of genotypes to consider.

In our model, the log-parameters are normally distributed and the natural parameters, which have a biological meaning, are log-normally distributed. We apply the ANOVA on both the log-parameters and the natural parameters, but it is usually considered that ANOVA is rather insensitive to departure from the normal assumption as long as the observations have the same non-normal parent distribution with possibly different means [18].

Global Wald test

The data are analysed with the model including the gene effect, M_{mult}. The significance of the gene effect coefficient is assessed by the following statistic:

$$W = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^T V^{-1} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \quad (4)$$

where V is the block for β_1 and β_2 of the estimation variance matrix. The statistic W is compared to the critical value of a χ^2 with 2 degrees of freedom.

Likelihood ratio test (LRT)

The data are analysed with M_{base} and M_{mult}. These two models are nested, thus the LRT can be used. The test statistic $-2 \times (L_{base} - L_{mult})$, where L_{base} and L_{mult} are the loglikelihood of respectively M_{base} and M_{mult}, is compared to the critical value of a χ^2 with 2 degrees of freedom, corresponding to the difference in the number of population parameters between the two models.

Study designs

We simulated data according to four designs. The first three have the same total number of observations and represent different trade-offs between the sample size N and the number of samples per patient n. The fourth design contains more subjects with many observations per patient to be closer to asymptotic conditions. Figure 1 illustrates the differences between the four designs regarding the samples allocation in time and the sampling size. The graph is composed of four rows (one per design) on top of the pharmacokinetic profile. Within each design, the sampling times of a

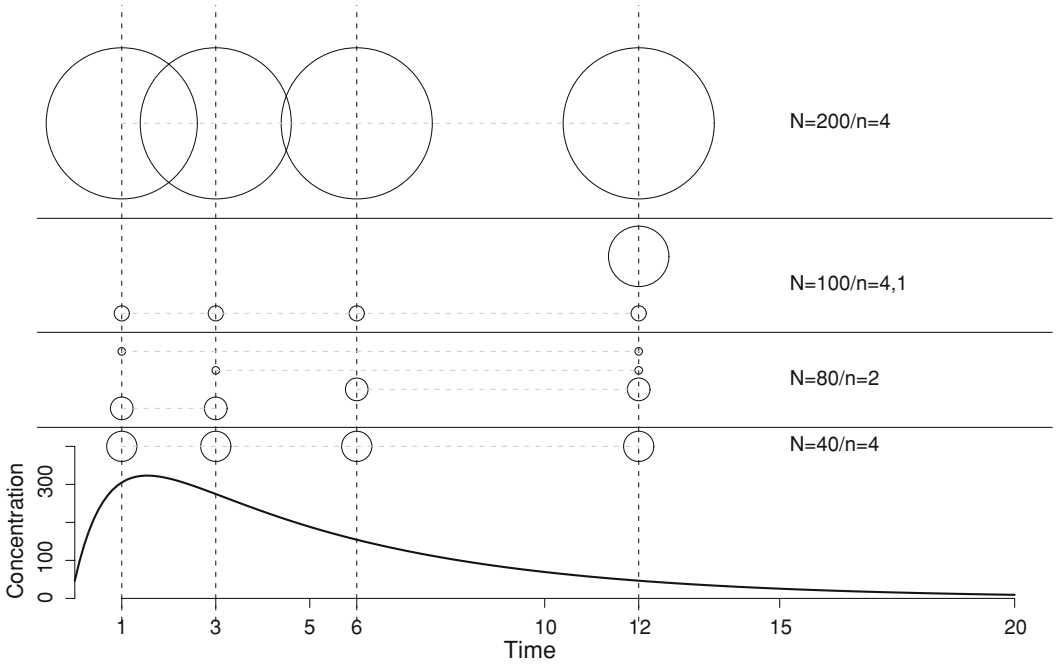


Fig. 1 Mean simulated concentration-time curve and allocation of the sampling times within each of the designs N = 40/n = 4, N = 80/n = 2, N = 100/n = 4,1 and N = 200/n = 4 (*separated by solid horizontal lines*): the vertical lines denote the four possible sampling times, the dashed horizontal lines join samples within the same group and the circles size is proportional to the sample size within each elementary design

group are represented as linked circles of size proportional to the number of subjects in the group with this sampling time.

(1) N = 40/n = 4
The first design is inspired from a real world example, the PK sub-study from the group of subjects receiving indinavir boosted with ritonavir b.i.d. in the COPHAR 2-ANRS 111 study, a multicentre non-comparative pilot trial of early therapeutic drug monitoring in HIV positive patients naïve of treatment [11, 12]. This design includes 40 subjects with 4 samples at time 1, 3, 6 and 12 h after the drug intake, which leads to a total of 160 observations. At the time of the study, these sampling times were empirically determined.

(2) N = 80/n = 2
In the second design, we require 80 subjects with two samples per patient and sampling times within the set of the original design. We used the Federov-Wynn algorithm that maximizes the determinant of the Fisher information matrix within a finite set of possible designs and which is implemented in the PFIM Interface 2.1 software [13]. We had to set the regression function *f*, the error model and a priori values of the population parameters (see Simulation study) as well as an initial guess for the population design. Regarding these constraints, the optimal design consists of 80 subjects sorted in two groups of 30 and two groups of 10 with two samples per subject respectively scheduled at 1 and 3 h, 6 and 12 h, 3 and 12 h and 1 and 12 h.
This configuration provides a rather sparse design keeping a total number of observations of 160.

(3) N = 100/n = 4,1
Third, we consider a pragmatic design with 20 subjects with the original set of sampling times (1, 3, 6 and 12 h) and 80 subjects with only a trough concentration (12 h) potentially collected in clinical routine. This combined design also contains a total number of observations of 160.

(4) N = 200/n = 4
The last design includes 200 subjects having the original set of sampling times.

Simulation study

The model and parameters used for the pharmacokinetic settings come from a preliminary analysis without covariates of the indinavir data described above using the FO algorithm implemented in NONMEM (see details in [8]). The concentrations are simulated using a one compartment model at steady state with first order absorption (k_a), first order elimination (k), a diagonal matrix for the random effects and a proportional error model (a fixed to 0). The dose is set to 400 mg. The fixed effects are $k_a = 1.4\ h^{-1}$, the apparent volume of distribution V/F = 102 l and $k = 0.2\ h^{-1}$, this parameterization was chosen to have only one parameter linked to the bioavailability, F. The between subjects variabilities on these parameters are respectively set to 113%, 41.3% and 26.4%. The coefficient of variation for the residual error is set to 20% (a = 0, b = 0.2). The first value in a series of simulated

concentration below the limit of quantification (LOQ = 0.02mg/l, according to the indinavir measurement technique in the COPHAR2 trial) is set to LOQ/2 and the remaining values are discarded [19].

The genetic framework is inspired from two SNPs of the ABCB1 gene coding for the P-glycoprotein, found to have an influence on the PK of protease inhibitors [20, 21]. We simulate a diplotype of SNP_1 and SNP_2 with C and G respectively the wild-type allele for the 2 exons and T the mutant allele. Their distribution mimic that of exon 26 and exon 21 of the ABCB1 gene as reported by Sakaeda et al. [22] yielding for SNP_1 unbalanced frequencies of 24%, 48% and 28% respectively for CC, CT and TT genotypes. As in the intestine, the P-glycoprotein restricts drug entry into the body we consider an effect on the drug bioavailability through the volume of distribution V/F, so that:

$$V/F_i = V/Fe^{\beta_{G1_i}} e^{\delta_{G2_i}} e^{\eta_{V/F,i}} \tag{5}$$

where $G1_i$ denotes the genotype for SNP_1 and $G2_i$ the genotype for SNP_2, β_{G1_i} is 0, β_1 or β_2 if $G1_i$ = CC, CT or TT and $\delta\delta G2_i$ is 0, δ_1 or δ_2 if $G2_i$ = GG, GT or TT. Under the null hypothesis both $e^{\beta_{G1_i}}$ and $e^{\delta_{G2_i}} = 1, 1, 1,$ whereas under the alternative hypothesis, we set a genetic model of co-dominance and multiplicative effects: $e^{\beta_{G1_i}} = 1, 1.2, 1.6$ and $e^{\delta_{G2_i}} = 1, 1.1, 1.3$. These values were chosen to be consistent with results found in the literature for ABCB1 polymorphisms on drugs disposition [23] and provide clinically relevant effect, with V/F and CL/F (= $k \times V/F$) increasing from 105.4 to 200.5 l and 21.1 to 40.1 l/h respectively between wild and mutant homozygotes for SNP_1. In the following, tests focus on the effect of SNP_1 even if we simulated diplotypes.

For the three designs (1), (2) and (3) with the same total number of observations, 1000 data sets are simulated both under the null (H_0) and the alternative hypothesis (H_1). The design (4) with N = 200/n = 4 is simulated only under H_0, providing evaluation of the type I error on 1000 data sets in conditions close to asymptotic to verify the convergence of the estimation algorithm. The technical description of the simulations is given in [8]. Figure 2 represents spaghetti plots of simulated concentrations versus time for the three designs with a total number of observations of 160, for one simulated data set respectively under H_0 and under H_1. According to their genotype for SNP_1 = CC, CT or TT, subjects curves are represented in plain, dashed or dotted lines, respectively, as well as the 12 h sample with circles, triangles or plus for subjects of the N = 100/n = 4,1 design. It is not readily apparent within each column which of the two data sets includes the gene effect.

Evaluation

In this work we use the SAEM algorithm implemented in the MONOLIX software version 2.1 [10]. The number of iterations during the two estimation phases and the number of Markov chains are set to provide fine convergence on one representative data set for each design under both hypotheses. Other parameters of the estimation algorithm are left to the default values.

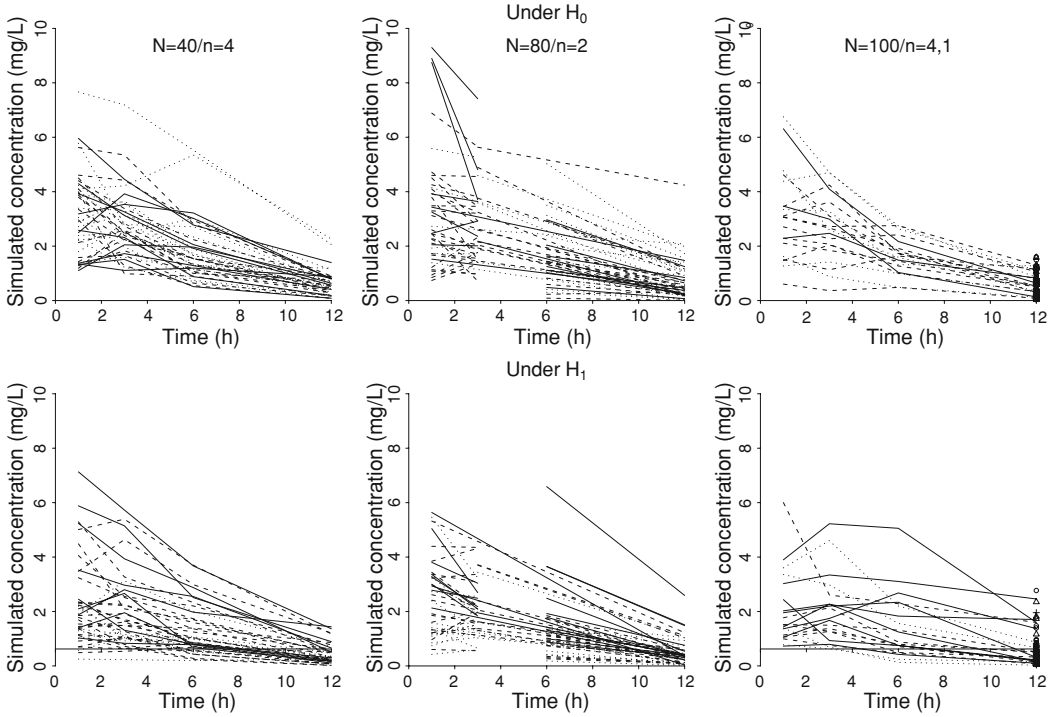


Fig. 2 Concentrations (ng/ml) simulated for the designs N = 40/n = 4 (left), N = 80/n = 2 (*center*) and N = 100/n = 4,1 (*right*) for a representative data set under H_0 (*top*) and a representative one under H_1 (*bottom*). Solid lines represent the subjects CC while dashed and dotted lines represent the subjects CT and TT for the exon SNP_1, respectively. For the N = 100/n = 4,1 design, circles represent the subjects CC while triangles and plus represent the subjects CT and TT for the exon SNP_1, respectively

On a given data set, the same seed is used to estimate parameters from M_{base} and M_{mult} but two different seeds are used for the importance sampling in the computation of the likelihood. A preliminary work was also performed to set the number of samples T of this importance sampling for each design. We considered 6 different values of T = 1000, 3000, 5000, 7000, 10000, 15000, 300000. For each value of T, the log-likelihood was estimated 25 times on one representative data set with both M_{base} and M_{mult} and the corresponding LRT was computed. The 25 estimations allowed us to discard any bias related to the choice of a seed as we used 5 different seeds for the random number generator at the estimation step and 5 different seeds for the random number generator at the importance sampling step. In the rest of the study, the number of samples T was set to a value that provides both a relative standard deviation on the 25 LRT estimates below 15% and moderate computing times.

Our work aims to evaluate the tests for the different designs dealing with statistical significance issues, which not necessarily imply clinical relevance [24]. First, the three tests are used to detect an effect of the SNP_1 (the effect of SNP_2 is not included in these analyses) on the bioavailability through the apparent volume of distribution parameter (V/F) in the 1000 data sets simulated under H_0 for the four designs. Then, the type I error of each test is computed as the percentage of data sets where the corresponding test was significant. Based on the central limit theorem and

with 5% the expectation for this percentage under H_0 the predicted interval around the type I error estimate is $[0.05 \pm 1.96 \times \sqrt{\frac{0.05\times(1-0.05)}{1000}}] = [3.6; 6.4]$. To ensure a type I error of 5%, we define a correction threshold as the 5th percentile of the distribution of the p-values of the test under H_0.

In a second step, for the designs N = 40/n = 4, N = 80/n = 2 and N = 100/n = 4,1 the tests are performed using the 1000 data sets simulated under H_1. Then, the power is defined as the percentage of data sets where the corresponding test was significant. We use the corrected threshold to compute the corrected powers, to allow comparison of the different tests taking into account the type I error different from 5%. In a third step, we have computed the data sets simulated under H_1 where the test was significant and at least one of the gene effect coefficient estimates (the absolute value) was clinically relevant i.e. greater than 20%. This calculation provided us with an estimate of each test ability to detect a clinically relevant effect on V/F (and thus CL/F) [24]. For the ANOVA only, one data set under H_0 and two data sets under H_1 where the number of subjects with a given SNP_1 was less than 2 were discarded from the analysis.

The ANOVA is based on the EBE for the parameter of interest, here the volume of distribution *V/F*. To assess the quality of the individual estimates from M_{base}, we compute the extent of the shrinkage on *V/F* for the four designs. A measure of the shrinkage of empirical Bayes estimates has been proposed by Savic et al. as 1 minus the ratio of the empirical standard deviation of η over the estimated standard deviation of the corresponding random effect [25]. Shrinkage estimators in literature are computed with a ratio of variances shrinking the observation toward the common mean [26, 27]. By analogy with these shrinkage estimators, in the present work, we define shrinkage on *V/F* as:

$$Sh_{\eta_{V/F}} = 1 - \frac{var\left(\eta_{V/F,i}\right)}{\omega^2_{V/F}} \tag{6}$$

where $var(\eta_{V/F,i})$ is the empirical variance of η for the volume of distribution and $\omega^2_{V/F}$ is the estimated variance of the corresponding random effect. A shrinkage, computed on standard deviation, over 30% is considered to potentially impact on covariates testing according to [25], therefore here we consider a threshold of 50%.

We also compare the empirical SE and the distribution of the SE obtained with SAEM for β_1 and β_2 for the different designs under both hypotheses. The empirical SE is defined as the sample estimate of the standard deviation from the β_1, β_2 estimates respectively on the 1000 simulated data sets.

To address point estimate and bias and how it may impact on the tests type I error and power, we compute the relative bias and relative root mean square error (RMSE) for V/F, $\omega^2_{V/F}$ and the residual error parameter b from M_{base} on the data sets simulated under H_0 and V/F, β_1, β_2, $\omega^2_{V/F}$ and b from M_{mult} on the data sets simulated under H_1. In addition, we have computed the relative bias and relative RMSE on the estimates obtained with FOCE-I in [8] on the N = 40/n = 4 and N = 200/n = 4 designs.

Results

The number of samples for the importance sampling, T, was set to 10000 and 15000 for the designs N = 40/n = 4 and N = 80/n = 2 and 20000 for both designs N = 100/n = 4,1 and N = 200/n = 4. SAEM achieves convergence on all data sets simulated with the four designs and each hypothesis.

Table 1 reports the estimated type I error for the three tests performed on the four designs. ANOVA has a correct type I error estimate for all designs with a value for the design at N = 80/n = 2 although close to the upper boundary. The results are analogous whether we consider the log-parameters or the natural parameters of the apparent volume of distribution (V/F), 5.5% and 5.3% respectively on the original design. The Wald test and the LRT, which are asymptotic tests, have significantly increased type I error in the three designs with a total number of observations equal to 160. Yet, the inflation remains moderate as all the estimates are below 10%. On the N = 200/n = 4 design, the Wald test and the LRT type I error returns to the nominal level of 5%.

The estimates for the power and the corrected power are given in Table 2, for the three designs N = 40/n = 4, N = 80/n = 2 and N = 100/n = 4,1. Before the correction, the Wald test and the LRT appeared wrongly more powerful than ANOVA. The ability to detect a clinically relevant effect is lower than the power to detect a statistically significant effect for the ANOVA, but identical for the Wald test and the LRT. In the following, we consider only the corrected power for comparisons across tests and designs as it accounts for the type I error inflation (or

Table 1 Type I error estimates (for 5% level test) on the N = 40/n = 4, N = 80/n = 2, N = 100/n = 4,1 and N = 200/n = 4 designs for each of the three tests using 1000 replicated data sets

		N = 40/n = 4	N = 80/n = 2	N = 100/n = 4,1	N = 200/n = 4
ANOVA	Log-parameters	5.5	6.2	3.8	4.2
	Natural parameters	5.3	6.4	4.3	5.0
Wald		8.9*	8.7*	8.4*	5.1
LRT		7.6*	7.8*	6.8*	5.9

* Outside the prediction interval for $5\% = [3.6 - 6.4]$

Table 2 Power estimates without and with ($Power_{corr}$) correction for the type I error inflation under H_0 on the N = 40/n = 4, N = 80/n = 2 and N = 100/n = 4,1 designs for each of the three tests using 1000 replicated data sets

		N = 40/n = 4		N = 80/n = 2		N = 100/n = 4,1	
		Power	$Power_{corr}$	Power	$Power_{corr}$	Power	$Power_{corr}$
ANOVA	Log-parameters	75.6	74.2	93.6	92.5	80.8	82.2
	Natural parameters	71.1	70.9	93.4	91.5	78.3	79.5
Wald		81.8	73.0	95.5	92.5	85.7	81.8
LRT		78.6	73.3	94.6	92.2	82.9	79.7

reduction for the ANOVA). For each design, the corrected power is rather analogous for the three tests within each design. For the three tests, the corrected power is greater for the design optimized using PFIM, with more subjects and less sample per subjects. In classical analysis increasing N improves the power and this also applies in longitudinal data analysis up to a point. Not only must N increase, but n also should be considered as well as the sampling schedule. This trade-off was achieved through optimal design and led to a satisfactory sparse design that even ANOVA, based on EBE, can handle.

Figure 3a displays the shrinkage for the apparent volume of distribution estimated using M_{base} on data sets simulated under H_0 and H_1 for the four designs under study. In Figure 3b and c, the type I error of the ANOVA on the log-parameters is plotted versus the median shrinkage for V/F under H_0 and the power of ANOVA on the log-parameters is plotted versus the median shrinkage for V/F under H_1. The median shrinkage is lower than 40% for the design N = 200/n = 4 under H_0 and for the designs N = 40/n = 4 and N = 80/n = 2 under both hypotheses. Only the design with N = 100/n = 4,1 subjects shows shrinkage with a potential impact on covariates testing, i.e. greater than 50%. This high value of shrinkage is essentially due to the 80 subjects with one sample (median value of shrinkage around 75% for these subjects versus 21% for the other subjects with 4 samples in this design). Under the alternative hypothesis, we simulated a mixture of normals with similar variance but three different means for the individual

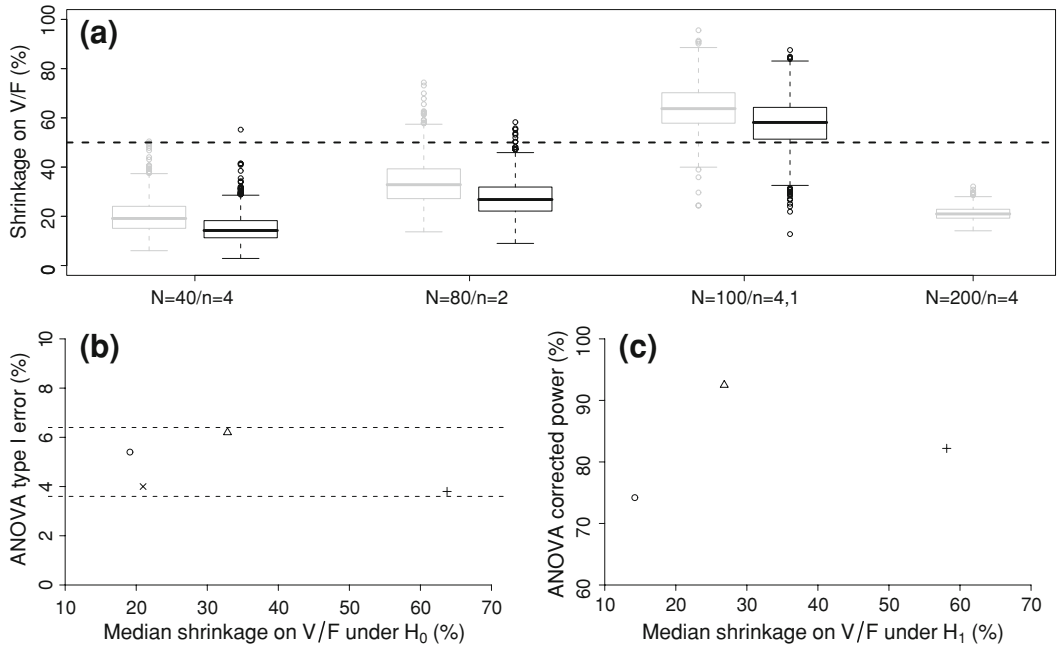


Fig. 3 **a** Boxplot of shrinkage on V/F from M_{base} obtained with SAEM on the 1000 data sets simulated under H_0 (*grey*) and H_1 (*black*) for the designs N = 40/n = 4, N = 80/n = 2, N = 100/n = 4,1 and N = 200/n = 4, **b** type I error for the ANOVA on the log-parameters versus the empirical shrinkage on V/F for the designs N = 40/n = 4 (◯), N = 80/n = 2 (Δ), N = 100/n = 4,1 (+) and N = 200/n = 4 (×) simulated under H_0, **c** Corrected power of the ANOVA on the log-parameters versus the empirical shrinkage on V/F for the designs N = 40/n = 4 (◯), N = 80/n = 2 (Δ) and N = 100/n = 4,1 (+) simulated under H_1

parameters of V/F. Under both hypothesis, the shrinkage is computed using the estimates from M_{base}. Under H_1, both the empirical variance of $\eta_{V/F,i}$ and the $\omega^2_{V/F}$ estimates are larger compared to the estimates under H_0. However, the empirical variance of $\eta_{V/F,i}$ increased more than $\omega^2_{V/F}$, thus the shrinkage estimates appeared to be consistently lower under H_1. For all designs under study, the type I error estimates of ANOVA remain within the prediction interval around 5% whereas the shrinkage estimates range from 19% to 64%. We do not observe a clear relationship between the power of ANOVA and the shrinkage on V/F, but the power decreases between the sparse and the combined design. Indeed, the ANOVA obtains a corrected power of 58% when performed only on the 80 subjects with one sample from the combined design, while on the optimized design with the same N but n = 2 its power was of 92.5%.

The relative Bias and RMSE for the estimated parameters are displayed in Table 3. SAEM and FOCE-I obtained unbiased estimates on both designs and similar relative RMSE except for V/F on the N = 200/n = 4 design where the

Table 3 Relative Bias and root mean square error (RMSE) in % evaluated from 1000 simulated data sets with M_{base} under H_0 for the volume of distribution (V/F), its interindividual variance ($\omega^2{}_{V/F}$) and the residual error parameter (b) for the N = 40/n = 4, N = 80/n = 2, N = 100/n = 4,1 and N = 200/n = 4 designs and from 1000 simulated data sets with M_{mult} under H_1 for V/F, β_1, β_2, $\omega^2_{V/F}$ and b for the N = 40/n = 4, N = 80/n = 2 and N = 100/n = 4,1 designs, using estimates from SAEM and FOCE-I when available in [8]

	Parameter	N = 40/n = 4		N = 80/n = 2	N = 100/n = 4,1	N = 200/n = 4	
		SAEM	FOCE-I	SAEM	SAEM	SAEM	FOCE-I
M_{base} **under** H_0							
Biais (%)	V/F	0.23	2.9	0.04	0.62	0.08	1.4
	$\omega^2_{V/F}$	−2.8	−0.6	0.2	−4.2	−0.8	0.7
	b	−0.3	−1.9	−3.8	−0.9	0.008	−1.8
RMSE (%)	V/F	8.6	9.5	8.5	11.8	3.8	11.1
	$\omega_{V/F}$	28.1	28.9	27.8	38.5	13.4	13.3
	b	8.8	10.3	15.8	12.4	4.0	4.8
M_{mult} **under** H_1							
Biais (%)	V/F	4.1	6.7	3.9	5.1		
	β_1	−1.0	−0.8	−2.5	−1.4		
	β_2	−1.0	−1.0	−1.8	−1.3		
	$\omega_{V/F}$	−7.5	−5.2	−1.3	−7.1		
	b	−0.6	−2.2	−3.5	−0.02		
RMSE (%)	V/F	17.9	19.2	15.0	19.9		
	β_1	19.9	20.0	15.3	18.1		
	β_2	21.7	21.7	16.5	21.3		
	$\omega_{V/F}$	29.7	29.6	26.7	39.1		
	b	9.22	10.1	16.8	13.0		

With FOCE-I, convergence was achieved and thus estimates were obtained from 969 and 950 data sets under H_0 and H_1 respectively for N = 40/n = 4 and 978 data sets under H_0 for N = 200/n = 4

expected improvement was observed only with SAEM. As the bias were null the discrepancies in RMSE across the designs arised only from the precision of estimation and the SE predicted by PFIM matched the lowest RMSE. Regarding the precision of estimation on β_1 and β_2 under both hypotheses for the designs under study in Fig. 4a, the SAEM algorithm shows good statistical properties: as expected, lower SE are observed for the design closer to asymptotic and the SE obtained with SAEM are close to their empirical value, albeit lightly under-estimated. Among the three designs with a total of 160 observations, the design N = 80/n = 2 provided the best performances; i.e., its empirical SE for estimates of the gene effect coefficients are the lowest. In Fig. 4b, the type I error of the Wald test is plotted versus the ratio of the median SE over the empirical SE for β_2 estimated under H_0. The under-estimation of the SE appears to be related to the type I error inflation of the Wald test as the three designs with a ratio below 0.98 have type I error estimates significantly above the nominal level. In Fig. 4c, the corrected power of the Wald test is plotted versus the empirical SE for β_2 estimated under H_1. The SE appears to be related to the power of the Wald test as it decreases as the SE increases with the highest power for the N = 80/n = 2 design.

Figure 5 represents the density function of a χ^2 with 2 degrees of freedom along with a focus on the values above 5.99 (the theoretical threshold) overlaid on a histogram of the LRT statistics obtained with the four designs simulated under H_0.

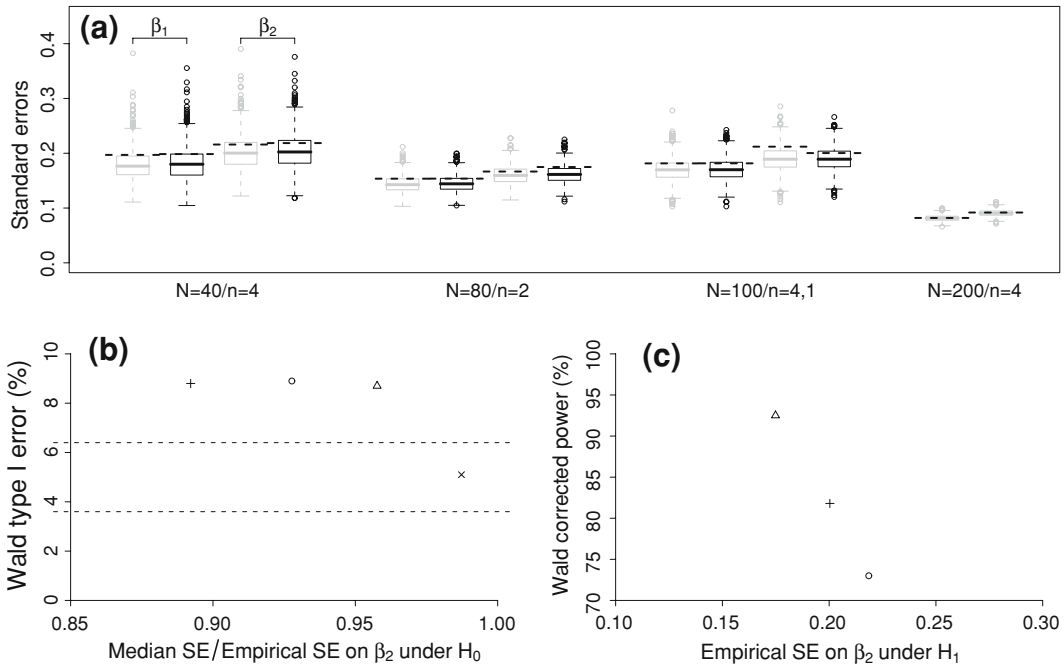


Fig. 4 **a** Boxplot of the estimated standard errors (SE) and corresponding empirical SE (*dotted line*) obtained with SAEM for β_1 and β_2 on the 1000 data sets simulated under both H_0 (*grey*) and H_1 (*black*) for the N = 40/n = 4, N = 80/n = 2, N = 100/n = 4,1 and N = 200/n = 4 designs, **b** Wald test type I error versus the ratio of the median SE over the empirical SE for β_2 for the designs N = 40/n = 4 (◯), N = 80/n = 2 (Δ), N = 100/n = 4,1 (+) and N = 200/n = 4 (×) simulated under H0, **c** Wald test corrected power versus the empirical SE for β_2 for the designs N = 40/n = 4 (◯), N = 80/n = 2 (Δ) and N = 100/n = 4,1 (+) simulated under H1

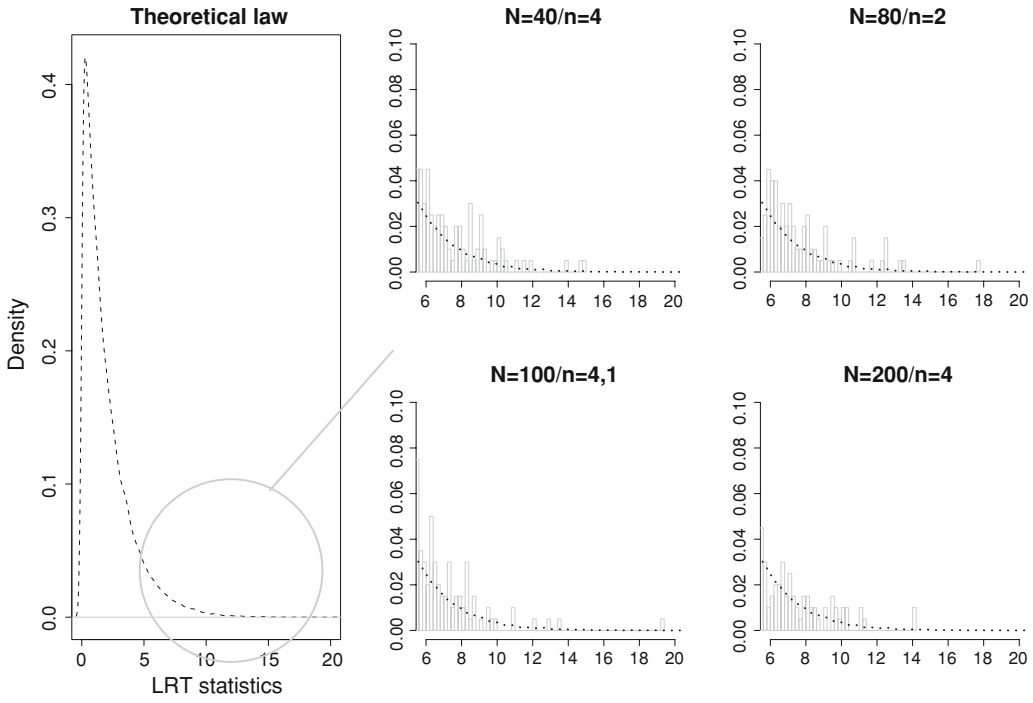


Fig. 5 Histograms of the likelihood ratio test (LRT) statistics above the theoretical threshold (5.99) obtained with SAEM under H_0 for the N = 40/n = 4, N = 80/n = 2, N = 100/n = 4,1 and N = 200/n = 4 designs. The *dotted curve* corresponds to the density of a χ^2 with 2 degrees of freedom

For the first three designs, the density curve is slightly shifted to the left compared to the histogram obtained under H_0 while for the N = 200/n = 4 design the superposition is complete.

Here, the corrected power of the Wald test is about 70% for the design N = 40/n = 4. In our previous work, we used the FOCE-I algorithm implemented in NONMEM version V [9] and we observed, for this design, a much lower corrected power of the Wald test (25%). Figure 6 displays, the standard errors of the gene effect coefficients β_1 (left) and β_2 (right) versus their estimates when using FOCE-I (top) or SAEM (bottom). With the FOCE-I algorithm, we observe a correlation between the estimate of the gene effect coefficients and its estimation error, that we do not observe with the SAEM algorithm. Such relationship leads to decreased values of the Wald statistic and therefore reduces the power to detect a gene effect.

Discussion

In the present study, we describe the impact of four designs on the performances of three tests for a pharmacogenetic effect in NLMEM using an exact maximum likelihood approach, the SAEM algorithm.

This work follows a previous study [8] which evaluated those three tests on two designs (N = 40/n = 4 and N = 200/n = 4) using the estimation algorithms FO and FOCE-I in NONMEM version V [9]. Type I error and power of Tables 1 and 2 in [8] can be compared to those in Tables 1 and 2 of the present paper respectively

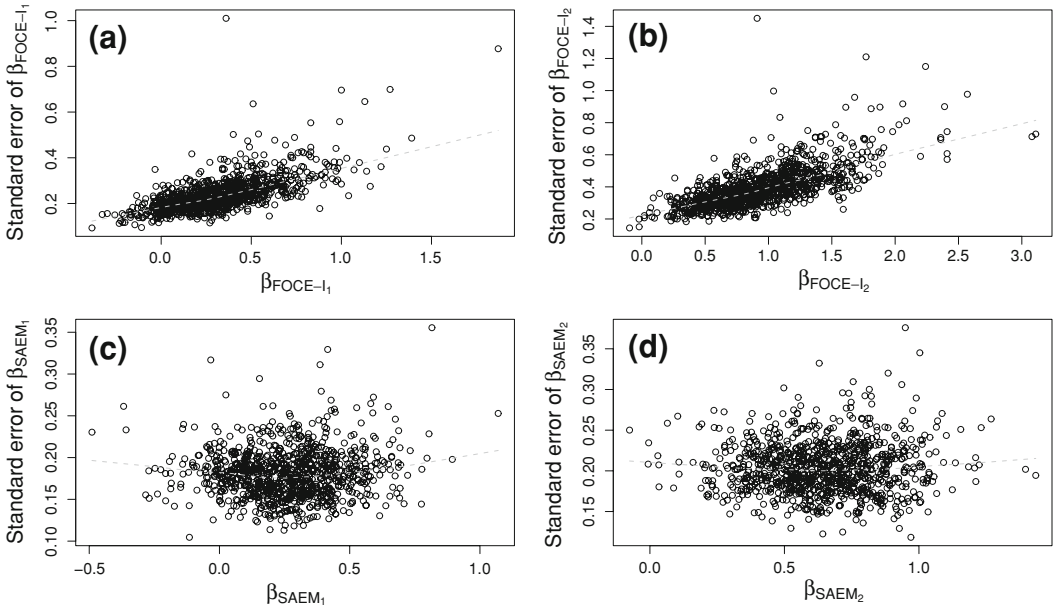


Fig. 6 Standard errors versus the estimates for β_1 and β_2 obtained with FOCE-I in NONMEM version V (**a**) and (**b**) and SAEM in MONOLIX version 2.1 (**c**) and (**d**) for the design N = 40/n = 4 simulated under H_1. Note that β_{FOCE-I_1} and β_{FOCE-I_2} correspond respectively to $e^{\beta_{SAEM_1}}$ and $e^{\beta_{SAEM_2}}$, therefore the scales are different

for the designs N = 40/n = 4 and N = 200/n = 4. The ANOVA in [8] was performed on the natural parameters. That simulation study has shown poor performances with the FO algorithm. The results obtained here with SAEM, in terms of type I error and power are rather similar to those obtained previously using FOCE-I, except for the Wald test. Indeed, with FOCE-I the type I error of the Wald test was still inflated on the design N = 200/n = 4 and the power was much lower. We hypothesised that the reduced power of the Wald approach could result partly from a poor estimation of the estimation variance matrix of the fixed effects due to the log-likelihood function linearisation, as we observed with FOCE-I a high correlation between the estimate and its estimation error. We did not meet this problem with SAEM. Besides, both algorithms obtained unbiased estimates with a similar improvement in relative RMSE on design N = 200/n = 4 except for V/F with FOCE-I. Moreover, FOCE-I had convergence problems for several data sets or did not provide the estimation variance matrix on design N = 40/n = 4 under H_1, while SAEM achieved convergence on all data sets whatever the design with the estimation variance matrix always provided. In the evaluation of model selection strategies in [8], we underlined the very poor performance of the Akaike criteria (AIC). This finding remains with SAEM (data not shown).

Other studies have evaluated by simulation the performance of tests for discrete covariate on continuous responses using NLMEM with various designs and estimation methods. The articles reporting these studies are summarized and sorted by year of publication in Table 4. Linearization based algorithms were mostly used with the exception of two recent works also using SAEM [17, 28]. Furthermore, categorical covariates were always simulated in two classes, apart from one study where it was up to three classes [29] and one study with continuous covariate [30].

In the present study, the ANOVA obtains the best performances with respect to type I error as no inflation is observed on the four designs, so there is no need in practice to correct the threshold for the test based on the EBE. This finding is in accordance with the results from Bonate et al. [31]. Considering t-tests on individual estimates, Comets et al. observed no inflation either [32]. Panhard et al. [33] obtained inflated type I error for t-tests for small n, however they studied cross-over trials where the model is fitted for each treatment separately and then the EBE are derived. With small n, the individual parameters estimates are thus shrunk toward the mean within each group, artificially increasing the statistic of the test. Analysing the whole data set, we thought that the ANOVA would be conservative in presence of sparse data, because shrinkage leads to regression of the individual parameters estimates towards the mean. Indeed, this phenomenon appears likely to reduce the test ability to discriminate means between the genotypes. In our study, the shrinkage may not have been strong enough as the sparse design was an optimal design and the one with more shrinkage had some subjects with rich design. Another advantage of ANOVA is that it requires only the model with no covariate to converge. It is noteworthy though that with unequal sample size within groups ANOVA is sensitive to heterogeneity of variances [34], this feature has not been studied in this simulation setting.

We explain the type I error inflation observed for the Wald test and the LRT by the designs with a total of 160 observations being far from the asymptotic. This result differ from those of Panhard et al. [33], Gobburu et al. [30] and Wählby et al. [29] which had similar trade-off in N an n given the number of model parameters with less that 160 observations (Table 3) as well as similar interindividual variability for the parameter of interest ($\approx$30%) and residual error variability (20–10%). Besides, Samson et al. [17] and Panhard et al. [28] observed no inflation of the type I error for these tests using SAEM for a covariate simulated in two classes with equivalent group size and at least $n = 6$. We hypothese therefore that the departure from the asymptotic found here is related to the covariate distribution, with only 11 mutant homozygotes in average for the design $N = 40/n = 4$. Distribution of genetic covariate (from a biallelic SNP with C and T, the wild and the mutant allele) is indeed very specific; the Hardy–Weinberg proportions [35] lead to proportions of 1/4, 1/2, 1/4 for CC, CT, TT being the less unbalanced of the possible distributions. Thus, we recommend to correct the type I error of asymptotic tests for genetic polymorphism with unbalanced genotypes including small number of subjects. Furthermore, such recommendation is relevant for any other covariate with several classes and very unbalanced distribution, such as disease status or tumor classes.

For the Wald test, we relate this inflation to the under-estimation of the SE of the gene effect coefficients. Indeed, when we performed the Wald test using the empirical SE rather than the estimated SE, we observed that the type I error was then no longer significantly different from the nominal level for all designs. Panhard et al. [36] observe this relationship with FOCE-I as well and show that modelling interoccasion variability in cross-over trials leads to a better estimation of the SE of the covariate effect coefficients providing type I errors of the Wald test and the LRT close to the nominal level. Here, the SE are obtained by MONOLIX after the

Table 4 Published simulation studies evaluating the performances of tests on discrete covariate effect for differents designs using nonlinear mixed effects models

First author (References)	Year	Algorithm	Software	Number of PK parameters	Design	Test	Covariate		Type I error
							Distribution (%)	Effect	
Bertrand (present study)	2009	SAEM	MONOLIX	3	N = {40,200}/ n = 4	ANOVA	24:48:28	1:1.2:1.6	No inflation for ANOVA
					N = 80/n = 2	Wald			Slight inflation for Wald and LRT when N = {40,80,100} corrected on N = 200
					N = 100/n = 4,1	LRT			
Panhard[a] [28]	2009	SAEM	MONOLIX	3	N = {40,24}/n = 10	Wald	50:50	–	No inflation
						LRT			
Bertrand [8]	2008	FO	NONMEM	3	N = {40,200}/n = 4	ANOVA	24:48:28	1:1.2:1.6	No inflation for ANOVA
		FOCE-I				Wald			Strong inflation for Wald and LRT with FO when N = {40,200}
						LRT			Slight inflation for Wald and LRT when N = 40 corrected on N = 200 with FOCE-I
Samson [17]	2007	SAEM	MONOLIX	4	N = {40,80, 200}/n = 6	Wald	50:50	1:{1.3,1.5}	No inflation
						LRT			
Panhard[a] [36, 33]	2007/2005	FOCE-I	R(nlme)	3	N = 12/n = 10	Student	50:50	1:{0.8,0.875, 0.9,1.1, 1.125,1.25}	Inflation for Student and Wilcoxon when n = 3, but not for n = {5,10}
					N = 24/n = 5	Wilcoxon			Inflation for Wald and LRT when N = {24,12}, but not for N = 40
					N = 40/n = 3	Wald			No inflation for Wald and LRT when modelling IOV
					N = {24,40,60}/ n = 10	LRT			

Table 4 continued

First author (References)	Year	Algorithm	Software	Number of PK parameters	Design	Test	Covariate		Type I error
							Distribution (%)	Effect	
Bonate [31]	2005	FOCE	NONMEM	2	All combinations of N = {50, 100,150,200}/ n = {2,4,6}	ANOVA	50:50	1:1.25	No inflation for ANOVA
						LRT			Inflation for LRT when n = {4,6}
Zhang [45]	2003	FOCE-I	NONMEM	4	N = 30/n = 5	LRT	50:50	–	No inflation
Gobburu [30]	2002	FO	NONMEM	3	N = 30/{n = 5,2,(5,2)}	LRT	Continuous	–	Strong inflation with FO
		FOCE							Sligh inflation with FOCE when n = {5,(5,2)}
		FOCE-I							No inflation with FOCE-I
Comets [32]	2001	FOCE	NONMEM	3	N = 20/n = 7	Wilcoxon	50:50	1:1.2	No inflation for Wilcoxon (individual fits)
		FOCE-I				LRT			Inflation for LRT with FOCE
									No inflation for LRT with FOCE-I
Lee [40]	2001	FOCE-I	R (nlme)	3	N = 200/n = 2	Student	90:10	1:1.3	Inflation for Student when n = 5,2
					N = 100/n = {2,3,(5,2)}	LRT	80:20		Inflation for LRT when n = {2,(5,2)}
							70:30		Inflation increases when the proportion of the subpopulation increases
							60:40		
							50:50		

Table 4 continued

First author (References)	Year	Algorithm	Software	Number of PK parameters	Design	Test	Covariate		Type I error
							Distribution (%)	Effect	
Wählby [29]	2001	FO	NONMEM	2	All combinations of N = {10,25,50, 250,1000}/n = {2,4,19}	LRT	98:2	–	Strong inflation with FOCE and Laplacian, when n = {4,19} and when N = {10,25,50}/n = 2 with addition of N = {250}/ n = 2 for FO
		FOCE					90:10		Slight inflation with -I method, when N = {10,25}
		FOCE-I					75:25		No impact of the covariate distribution, when N = 50/ n = 2
		Laplacian					33:33:33		
		Laplacian-I							
White [39]	1992	FO	NONMEM	2	All combinations of N = {60,75,100}[b] /n = {10,2}	Wald	85:15[4]	1:{0.9,0.8,0.7,0.6}	Inflation for both Wald and LRT on all designs
						LRT	70:30		Inflation increases when n and/or the proportionof the subpopulation increases
							50:50		

ANOVA, Student and Wilcoxon tests are based on empirical Bayes estimates of the individual parameters

[a] Cross-over trials

[b] The size of the control group remains 50, only the size of the comparison group varies

estimation with SAEM using a linearization of the model around the conditional expectation of the individual parameters, yet Dartois et al. [37] have also observed under-estimated SE when using the computation approach based on Louis' principle [38]. With SAEM, as expected, the inflation did not worsen when increasing the number of samples per subjects as reported for FO, FOCE-I in NONMEM [39, 29, 31, 30, 32] or FOCE-I in nlme [40, 33, 36]. This slight inflation can be handled using randomisation tests [41], computing the true distribution of the statistic for the data set under study and deriving a *p*-value. Approximate tests could also be used with degrees of freedom derived from the information in the design i.e. accounting for k, n and N [42], although there is no real consensus on how to do it for nonlinear mixed effect models. An additional advantage of the Wald test is that only the model including the covariate is required and, assuming symmetric confidence intervals, it is not a problem to test if the gene effect coefficients equal 0.

To assess the power, we have simulated a 60% increase in V/F which leads to a relevant adjustment in the dose in the TT genotypes for SNP_1; a 40% increase. There was no or slight changes in the proportion of data sets simulated under H_1 where the three tests were significant when considering for a clinically relevant genetic effect, with the exception of the ANOVA on the design N = 100/n = 4,1. We show the impact of the shrinkage due to the subjects with only one sample in the design N = 100/n = 4,1 on the ANOVA performance. In our simulation setting, the reduction in the test ability to discriminate means between the genotypes is more pronounced under the alternative hypothesis. For the N = 40/n = 4 design the median shrinkage was 14.2% for V/F (Fig. 3) and 30.5% and 36.2% for k_a and k respectively, thus the shrinkage should not have impacted on the power of the ANOVA had the effect been assessed on those parameters. Besides, the shrinkage was also found to be lower under the alternative hypothesis, further research on this trend would be interesting. For the Wald test, we show a direct relationship between the design, the precision of estimation for the covariate effect and the power. Indeed the design N = 80/n = 2 optimised using PFIM has both the lowest SE on the gene effect coefficients (β_1 and β_2) and the highest power. Our previous results with FOCE-I also underline that unbiased SE estimates are required to perform the Wald test. We should note however that we used the population model without covariate for design optimisation. Our results are in accordance with the work performed by Retout et al. [14]. Indeed, they studied design optimization to improve the power of the Wald test using a model including the covariate and also found that the power increases when the number of subjects increases and the number of samples per subject decreases. For this work, Retout et al. developed the Fisher information matrix for population model with covariate. But this development has not yet been implemented in the available version of the PFIM software. One extension of the present work would be to investigate other criteria such as D_S-optimality criterion to design pharmacogenetic studies specifically focusing on gene effect coefficients.

In the choice of the two additional designs compared to [8] used for this simulation study, we account for practical considerations. Basically, we increased the number of subjects to fit the requirements of the EMEA [3]. However, increasing the number of subjects can lead to practical issues in terms of blood sampling, as extensive sampling can not be performed in all subjects for practical

reasons. Therefore, we consider two designs. First, an exploratory study where we use PFIM to define different groups with two samples per subjects within a predefined set of sampling times. This approach could be used in studies with pharmacogenetics as primary endpoint when the population pharmacokinetic model is already known; for instance, studies on pharmacokinetic evaluation of a chemical entity when the genetic variation is likely to translate into important differences in the systemic exposure. Second, a more practical study in which we use trough concentrations collected during routine monitoring as well as a small group of subjects with more extensive sampling. The latter could be a phase III or IV clinical study where genotyping will support recommendations for use in genetic subpopulations [43].

In this work, we assume that the gene effect only acts on a single parameter, the bioavailability, so we use k (the elimination constant rate) rather than CL/F in order to have only one parameter related to F, the oral volume of distribution V/F. However, population models are more commonly parameterized using CL/F, thus another perspective of this work would be to consider a gene effect on several parameters: CL/F and V/F. Besides, more than one exon control the complex pathway leading from DNA to metabolic activity. Thus, it would be interesting to investigate how model-based tests handle haplotypes [44] which lead to a larger number of unbalanced classes. Here, we could hardly consider haplotypes due to the small sample sizes. Finally, investigating genes not on the same chromosome will also raise the issue of multiple covariates.

In conclusion, the ANOVA can be applied easily and performs satisfactorily as long as the design provides low shrinkage on the parameter of interest. Whereas for asymptotic tests, a correction has to be performed on designs with unbalanced genotypes including small number of subjects. Design optimization algorithms for models with covariate are well suited and offer perspectives to handle pharmacogenetic studies but have still to be implemented in the available softwares.

Acknowledgments We would like to thank the COPHAR 2-ANRS 111 scientific committee (investigators: Pr. D. Salmon and Dr. X. Duval, pharmacology: Pr. J.M. Tréluyer, methodology: Pr. F. Mentré) for giving us access to the pharmacogenetic data of the indinavir arm in order to build our simulations. We would also like to thank the IFR02 of INSERM and Hervé Le Nagard for the use of the "centre de biomodélisation" as well as the Pr. Marc lavielle for the precious help he provided in using MONOLIX. During this work, Céline M. Laffont was working at the Institut de Recherches Internationales Servier as pharmacometrician and Julie Bertrand was supported by a grant from the Institut de Recherches Internationales Servier.

References

1. EMEA (2008) ICH topic E15 definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories. Technical report, EMEA
2. FDA (2008) E15 definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories. Technical report, FDA
3. EMEA (2007) Reflection paper on the use of pharmacogenetics in the pharmacokinetic evaluation of medicinal products. Technical report, EMEA
4. Hu XP, Xu JM, Hu YM, Mei Q, Xu XH (2007) Effects of CYP2C19 genetic polymorphism on the pharmacokinetics and pharmacodynamics of omeprazole in chinese people. J Clin Pharmacol Ther 32:517–524

5. Hirt D, Mentré F, Tran A, Rey E, Auleley S, Salmon D, Duval X, Tréluyer JM, COPHAR2-ANRS Study Group (2008) Effect of CYP2C19 polymorphism on nelfinavir to M8 biotransformation in HIV patients. Br J Clin Pharmacol 65:548–57
6. Yamasaki Y, Ieiri I, Kusuhara H, Sasaki T, Kimura M, Tabuchi H, Ando Y, Irie S, Ware JA, Nakai Y, Higuchi S, Sugiyama Y (2008) Pharmacogenetic characterization of sulfasalazine disposition based on NAT2 and ABCG2 (Bcrp) gene polymorphisms in humans. Clin Pharmacol Ther 84:95–103
7. Li D, Lu W, Zhu JY, Gao J, Lou YQ, Zhang GL (2007) Population pharmacokinetics of tacrolimus and CYP3A5, MDR1 and IL-10 polymorphisms in adult liver transplant patients. Clin Pharm Ther 32:505–515
8. Bertrand J, Comets E, Mentré F (2008) Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. J Biopharm Stat 18: 1084–1102
9. Sheiner L, Beal S (1998) NONMEM Version 5.1. University of California, NONMEM Project Group, San Francisco
10. Lavielle M (2008) MONOLIX (MOdèles NOn LInéairesà effets miXtes). MONOLIX Group, Orsay, France, 2008. http://www.software.monolix.org/index.php
11. Duval X, Mentré F, Rey E, Auleley S, Peytavin G, Biour M, Métro A, Goujard C, Taburet AM, Lascoux C, Panhard X, Tréluyer JM, Salmon D (2009) Benefit of therapeutic drug monitoring of protease inhibitors in HIV-infected patients depends on PI used in HAART regimen—ANRS 111 trial. Fundam Clin Pharmacol. in press
12. Bertrand J, Treluyer JM, Panhard X, Tran A, Rey SE, Salmon-Céron D, Duval X, Mentré F, COPHAR2-ANRS 111 Study Group (2009) Influence of pharmacogenetics on indinavir disposition and short-term response in HIV patients initiating HAART. Eur J Clin Pharmacol 65:667–678
13. Retout S, Comets E, Le Nagard H, Bazzoli C, Mentré F (2007) PFIM Interface 2.1. UMR738, INSERM, Université Paris 7, Paris, France, http://www.pfim.biotstat.fr
14. Retout S, Comets E, Samson A, Mentré F (2007) Design in nonlinear mixed effects models: optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates. Stat Med 26:5162–5179
15. Pillai GC, Mentré F, Steimer JL (2005) Non-linear mixed effects modeling—from methodology and software development to driving implementation in drug development science. J Pharmacokinet Pharmacodyn 32:161–183
16. Deylon B, Lavielle M, Moulines E (1999) Convergence of a stochastic approximation version of EM algorithm. Ann Stat 27:94–128
17. Samson A, Lavielle M, Mentré F (2007) The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. Stat Med 26:4860–4875
18. Box GEP, Andersen SL (1955) Permutation theory in the derivation of robust criteria and the study of departures from assumption. J R Stat Soc Ser B Meth 17:1–34
19. Beal SL (2001) Ways to fit a PK model with some data below the quantification limit. J Pharmacokinet Pharmacodyn 28:481–504
20. Fellay J, Marzolini C, Meaden E, Back D, Buclin T, Chave J (2002) Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter gene MDR1: a pharmacogenetic study. Lancet 359:30–36
21. Solas C, Simon N, Drogoul NP, Quaranta S, Frixon-Marin V, Bourgarel-Rey V, Brunet C, Gastaut JA, Durand A, Lacarelle B, Poizot-Martin I (2007) Minimal effect of MDR1 and CYP3A5 genetic polymorphisms on the pharmacokinetics of indinavir in HIV-infected patients. Br J Clin Pharmacol 64:353–362
22. Sakaeda T, Nakamura T, Okumura K (2002) MDR1 genotype-related pharmacokinetics and pharmacodynamics. Biol Pharm Bull 25:1391–1400
23. Marzolini C, Paus E, Buclin T, Kim RB (2003) Polymorphisms in human MDR1 (p-glycoprotein): recent advances and clinical relevance. Clin Pharmacol Ther 75:13–33
24. Sackett DL, Haynes RB, Guyatt GH, Tugwellrada P (1991) Clinical epidemiology. A basic science for clinical medecine, 2nd edn. Little Brown, Boston.
25. Karlsson MO, Savic RM (2007) Diagnosing model diagnostics. Clin Pharm Ther 82:17–20
26. Robert CP (1994) The Bayesian Choice. A decision-theoretic motivation. Springer-Verlag, New York
27. Verbeke G, Molenberghs G (2000) Linear mixed models for longitudinal data. Springer-Verlag, New York
28. Panhard X, Samson A (2009) Extension of the SAEM algorithm for nonlinear mixed models with 2 levels of random effects. Biostatistics 10:121–135

29. Wählby U, Jonsson EN, Karlsson MO (2001) Assessment of actual significance levels for covariate effects in NONMEM. J Pharmacokinet Pharmacodyn 28:231–252
30. Gobburu JV, Lawrence J (2002) Application of resampling techniques to estimate exact significance levels for covariate selection during nonlinear mixed effects model building: some inferences. Pharm Res 19:92–98
31. Bonate PL (2005) Covariate detection in population pharmacokinetics using partially linear mixed effects models. Pharm Res 22:541–549
32. Comets E, Mentré F (2001) Evaluation of tests based on individual versus population modeling to compare dissolution curves. J Biopharm Stat 11:107–123
33. Panhard X, Mentré F (2005) Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence cross-over trials. Stat Med 24:1509–1524
34. Box GEP (1954) Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. Ann Math Stat 25:290–302
35. Crow JF (1999) Hardy, Weinberg and language impediments. Genetics 152:821–825
36. Panhard X, Taburet AM, Piketti C, Mentré F (2007) Impact of modelling intra-subject variability on tests based on non-linear mixed-effects models in cross-over pharmacokinetic trials with application to the interaction of tenofovir on atazanavir in HIV patients. Stat Med 26:1268–1284
37. Dartois C, Lemenuel-Diot A, Laveille C, Tranchand B, Tod M, Girard P (2007) Evaluation of uncertainty parameters estimated by different population pk software and methods. J Pharmacokinet Pharmacodyn 34:289–311
38. Kuhn E, Lavielle M (2005) Maximum likelihood estimation in nonlinear mixed effects models. Comput Stat Data Anal 49:1020–38
39. White DB, Walawander CA, Liu DY, Grasela TH (1992) Evaluation of hypothesis testing for comparing two populations using NONMEM analysis. J Pharmacokinet Biopharm 20:295–313
40. Lee PI (2001) Design and power of a population pharmacokinetic study. Pharm Res 18:75–82
41. Manly BFJ (1998) Randomization, bootstrap and Monte Carlo methods in biology, 2nd edn. Chapman & Hall, London
42. Elston DA (1998) Estimation of denominator degrees of freedom of F-distributions for assessing Wald statistics for fixed-effect factors in unbalanced mixed models. Biometrics 54:1085–1096
43. Chou M, Bertrand J, Segeral O, Verstuyft C, Borand L, Comets E, Becquemont L, Ouk V, Mentré F, Taburet AM (2009) Inter- and intra-patient variabilities in nevirapine plasma concentrations in HIV-infected cambodian patients and the effect of CYP2B6 genetic polymorphism: ANRS 12154 study. 16th conference on retroviruses and opportunistic infections, Montreal, Canada
44. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296: 2225–2229
45. Zhang L, Beal SL, Sheiner LB (2003) Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. J Pharmacokinet Pharmacodyn 30:387–404

2.3 Alternatives aux tests asymptotiques pour détecter un effet gène sur un paramètre pharmacocinétique

2.3.1 Résumé

Nos travaux précédents ont montré que seule l'ANOVA conservait une erreur de type I non significativement différente de 5% pour tester l'influence d'un polymorphisme génétique sur un paramètre pharmacocinétique, sur plusieurs plans d'expérience communément employés. Cependant, l'ANOVA ne permet pas d'évaluer certains types de modèle de l'effet gène (Arab-Alameddine et al., 2009) et reste essentiellement utilisée pour les analyses exploratoires. En effet, la grande majorité des études de pharmacogénétique en pharmacocinétique de population utilise le LRT. Dans ce contexte, nous avons voulu évaluer deux alternatives au test de Wald et au LRT sur les plans d'expérience avec peu de sujets et/ou peu de prélèvements par sujet.

La première alternative est une approche non-paramétrique où nous nous affranchissons complètement de la distribution de référence (χ^2) pour construire la distribution de la statistique sous l'hypothèse d'absence d'un effet gène à l'aide de permutation du vecteur de covariables de l'échantillon observé. La seconde alternative n'est appliquée que pour le test de Wald qui repose sur les erreurs d'estimation, et fait ainsi une approximation supplémentaire par rapport au LRT. Elle consiste à comparer la statistique obtenue à une distribution de Fisher dont le degré de liberté au dénominateur (df) permet de corriger l'éloignement à l'asymptotique. Nous avons considéré quatre approches différentes :

i. DF_{ANOVA} où le calcul du df est basé sur la décomposition classique dans l'ANOVA (Pinheiro et Bates, 2000)

ii. DF_{RAND} où le df est le nombre de sujets moins le nombre de variances estimées ; cette approche est mise en œuvre dans la PROC NLMIXED du logiciel SAS (SAS Institute Inc, 2004)

iii. DF_{MNLM} où le df est calculé comme la différence entre le nombre de sujets (N) et le nombre d'effets fixes estimés et une correction de la matrice de variance d'estimation est réalisée par un facteur N/df ; cette approche a été proposée dans les modèles non linéaires multivariés par Gallant (Gallant, 1975)

iv. DF_{FC} où le calcul du df repose sur la matrice de variance d'estimation ; nous proposons une extension aux MNLEM de la méthode proposée par Fai et Cornelius (1996) dans le cadre des modèles linéaires mixtes, en utilisant la matrice de Fisher approximée par linéarisation.

Ces alternatives ont été évaluées avec les algorithmes d'estimation SAEM et FO sur le plan d'expérience avec N=40 sujets. Nous avons utilisé le même cadre de simulation que celui présenté dans les parties 2.1 et 2.2.

Les tests de permutations s'avèrent une alternative valide à la correction par simulation, leur principal avantage étant de ne reposer que sur l'hypothèse d'interchangeabilité qui est respectée dans ce contexte. Leur principale limite reste le coût en temps de calcul qui dépend des performances de l'algorithme, et qui rend impératif l'utilisation d'un algorithme rapide et numériquement stable. Dans notre étude, seule la méthode DF_{MNLM} permet de corriger l'inflation de l'erreur de type I du test de Wald avec SAEM. Mais contrairement à l'approche par permutation, cette méthode ne permet pas de corriger l'inflation plus importante observée avec l'algorithme d'estimation FO. Le biais entrainé par la linéarisation réalisée dans l'algorithme FO est trop important pour que la pondération par N/df puisse corriger la sous-évaluation des erreurs d'estimation.

Les tests de permutation peuvent être aisément employés dans le contexte des études pharmacogénétiques en pharmacocinétique de population avec un algorithme exact tel que SAEM. La méthode DF_{MNLM} offre une alternative plus économique en temps de calcul, mais sa performance devrait être confirmée dans d'autres cadres de simulation.

Ces méthodes et résultats sont présentés dans un article en préparation pour le journal *Biometrics*.

2.3.2 Article 3 (en préparation)

Some Alternatives to Asymptotic Tests for the Analysis of Pharmacogenetic Data using Nonlinear Mixed Effects Models

Julie Bertrand[1,*], Emmanuelle Comets[1], Marylore Chenel[2] and France Mentré[1]

[1] UMR 738, INSERM, Université Paris Diderot, Paris F-75018, France

[2] Institut de Recherches Internationales Servier, Courbevoie F-92400, France

**email:* julie.bertrand@inserm.fr

SUMMARY: Nonlinear mixed effect models allow to investigate individual differences in drug concentration profiles (pharmacokinetics, PK) and responses. Pharmacogenetics focus on the genetic component of this variability. Two tests often used to detect a gene effect on a PK parameter are i) the Wald test, assessing whether estimates for the gene effect are significantly different from 0 and ii) the likelihood ratio test (LRT) comparing models with and without the genetic effect. A correction is required to account for the inflation of the type I error observed with those two tests on small sample size and/or with unevenly distributed genotypes. In this work, we develop two alternatives to these asymptotic tests and evaluate them by means of a simulation study. First, we assess the interest of permutation test using the Wald and the LRT statistics. Second for the Wald test we propose the use of four F-distribution based approaches with various values of denominator degrees of freedom. We also explore the influence of the estimation algorithm using both the linearisation-based algorithm First-Order approximation and the exact algorithm Stochastic Approximation Expectation Maximization (SAEM). We apply our results to the analysis of the pharmacogenetic of indinavir in HIV patients recruited in the COPHAR2-ANRS 111 trial. Results of the simulation study show that the permutation test seems appropriate with an exact estimation algorithm ensuring reasonable computing times and/or no numerical difficulties, like SAEM. One of the four F-distribution based approaches provides a correct type I error estimate for the Wald test and should be further investigated.

KEY WORDS: F-distribution based approach; First-order approximation; Nonlinear mixed effects models; Permutation tests; Pharmacogenetics.

1. Introduction

Pharmacokinetics (PK) studies the time course of a drug in the body. Nonlinear mixed effects models (NLMEM) in the analysis of PK data allow to integrate the knowledge accumulated on the drug Absorption, Distribution, Metabolism and Elimination (ADME) to quantify the interindividual variability with fewer samples per patient than standard non compartmental approach (Gabrielsson and Weiner, 1999). Pharmacogenetics (PG) studies the relationship between this interindividual variability and variations in DNA sequence of proteins involved in the ADME mechanisms of the drug.

In the analysis of repeated measurements, testing for differences between groups, for instance according to genotypes, is usually performed through one of three tests : the Wald test, the likelihood ratio test (LRT) and the Score test (Vonesh and Chinchilli, 1997). The latter is less often used in NLMEM as it is not implemented in the available softwares and has not been much evaluated (Guedj et al., 2007; Jacqmin-Gadda et al., 2009). In previous simulation studies, we have shown that a correction for type I error inflation was required for the Wald test and the LRT in data sets of small sample size and/or with unevenly distributed genotypes (Bertrand et al., 2008, 2009b).

The aim of the present work is to investigate two alternatives to the asymptotic tests to detect a gene effect in PG studies; permutation and F-distribution based tests. Permutation testing is a way of determining whether the null hypothesis of randomness is reasonable, i. e. whether the pattern present in the data is a purely chance effect of observations in a random order (Good, 1994; Manly, 1998). This alternative requires fewer assumptions than a correction based on simulations from the model under the null hypothesis, but is seldom used in NLMEM (Ding and Wu, 2001; Holford, 2001).

The second alternative is a correction for the Wald test, aiming to correct for the underestimation of the estimation variance terms on small sample size. We investigate four such

corrections based on F-distribution with various values of denominator degrees of freedom (df). The first approach uses a denominator df derived from the classical decomposition of df in balanced, multilevel one-way analysis of variance design (Pinheiro and Bates, 2000) and is implemented in the nlme function in R (R Development Core Team, 2008). The second approach was proposed for NLMEM by Wolfinger (2000) and is implemented in the NLMIXED Macro in SAS (SAS Institute Inc., 2004). In the third approach we propose an adaptation for NLMEM of a method developed by Gallant (1975) in multivariate nonlinear models (MNLM) which Vonesh and Chinchilli (1997) have also evaluated for generalized multivariate nonlinear models for repeated measurements. As last approach, we propose an extension to NLMEM of the method developed by Fai and Cornelius (1996) for linear mixed effect models (LMEM) that is implemented in the DDFM=SAT option of the MIXED Macro in SAS.

To our knowledge, these alternatives have not yet been evaluated and compared through a thorough simulation study in NLMEM. In the present work, we study their appropriateness through simulations. The permutation tests are evaluated for both the Wald and the LRT statistics whereas the use of F-distribution is applied for the Wald test only. The two alternatives are evaluated in terms of type I error and power using the same simulation setting as in our previous studies, in order to compare the results with simulation based corrections. The simulation setting is based on the indinavir data from a sub-study performed in the COPHAR2-ANRS 111 trial (Duval et al., 2009) involving 40 HIV-positive patients receiving indinavir/ritonavir from whom four samples were collected at 1, 3, 6 and 12h following administration. As an illustration, we present an analysis of the indinavir concentration-time profiles collected in the COPHAR2 trial in which we have used both alternatives given the small number of subjects and the unevenly distributed genotypes (Bertrand et al., 2009a).

As in NLMEM the integral in the likelihood has no analytical form, specific algorithms

are needed to estimate the model parameters and their standard error (SE) (Pillai et al., 2005). Some algorithms propose to use model linearisation like the first order (FO) (Sheiner et al., 1972) and first order conditional estimation with interaction (FOCE-I) (Lindstrom and Bates, 1990) methods implemented in the NONMEM software (Sheiner and Beal, 1998). Others consider a numerical approximation of the likelihood like the Laplacian or Adaptive Gaussian quadrature algorithms (AGQ) (Wolfinger, 1993), which are implemented in the NLMIXED Macro in SAS. The more recent stochastic approximation EM algorithm (SAEM) (Deylon et al., 1999), implemented in the MONOLIX software (Lavielle, 2008), uses MCMC methods and a stochastic version of the EM algorithm to obtain the maximum likelihood estimates of the NLMEM parameters without linearisation. The AGQ and SAEM algorithms are exact methods and show better performances than linearisation based algorithms (Girard and Mentré, 2005). In Bertrand et al. (2008, 2009b), we have shown that the bias in the FO algorithm leads to a very large inflation of both asymptotic tests type I error while with FOCE-I similar results as SAEM were obtained but we met difficulties achieving convergence and obtaining standard errors. Thus, to account for the influence of the estimation algorithm in the present work we use FO and SAEM. FOCE-I was not used, because of prohibitive computing times and numerical issues, especially for the evaluation of the permutation tests.

In section 2, we present the model and how the likelihood and the estimation variance matrix are obtained. Then, we introduce the usual asymptotic tests and both investigated alternatives in section 3. In sections 4 and 5 we describe the real data, the simulation study and the evaluation protocol. The results of the evaluation and the illustration are presented in section 6. We finally discussed our findings in section 7.

2. Evaluation of the likelihood and the estimation variance matrix in NLMEM

2.1 *Models and notation*

To describe the $(n_i \times 1)$-vector of concentrations $\boldsymbol{y_i}$ of a subject i, we use a pharmacokinetic function f nonlinear in its parameters $\boldsymbol{\phi_i}$:

$$\boldsymbol{y_i} = f(\boldsymbol{X_i}; \boldsymbol{\phi_i}) + \boldsymbol{\epsilon_i} \ , \tag{1}$$

where $\boldsymbol{X_i}$ is the within-subject design $(n_i \times 1)$-vector and $\boldsymbol{\phi_i} = \boldsymbol{\mu} + \boldsymbol{A_i}\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{\eta_i}$ is the subject $(p \times 1)$-vector of parameters. $\boldsymbol{A_i}$ is the $(p \times k)$-covariate matrix which permits to model the relationship between the covariates and $\boldsymbol{\phi_i}$ with $\boldsymbol{\beta}$ the corresponding $(k \times 1)$-coefficient effects vector. $\boldsymbol{\theta} = [\boldsymbol{\mu}' \quad \boldsymbol{\beta}']'$ is the $((p+k) \times 1)$-fixed effects vector. $\boldsymbol{B}$ is a $(p \times q)$-design matrix, permitting some components of $\boldsymbol{\phi_i}$ to have no associated random effect when $p > q$. $\boldsymbol{\eta_i}$ is the random effect $(q \times 1)$-vector which follows a Gaussian distribution with null mean and variance-covariance $(q \times q)$-matrix $\boldsymbol{\Omega}$. $\boldsymbol{\epsilon_i}$ is the residual error $(n_i \times 1)$-vector which follows a Gaussian distribution with null mean and variance-covariance $(n_i \times n_i)$-matrix $\boldsymbol{\Sigma_i}(\boldsymbol{X_i}; \boldsymbol{\phi_i}, \boldsymbol{\gamma}) = diag(g(\boldsymbol{X_i}; \boldsymbol{\phi_i}, \boldsymbol{\gamma})^2)$.

To obtain a combination of constant and proportional error models, $g(\boldsymbol{X_i}; \boldsymbol{\phi_i}, \boldsymbol{\gamma})$ can be set equal to $a + bf(\boldsymbol{X_i}; \boldsymbol{\phi_i})$ with $\boldsymbol{\gamma} = [a \quad b]'$ the vector of the error model parameters. Let define the $(l \times 1)$-vector of variance parameters as $\boldsymbol{\lambda} = [Vech(\boldsymbol{\Omega})' \quad \boldsymbol{\gamma}']'$ where the operator $Vech(.)$ creates a column vector from the matrix $\boldsymbol{\Omega}$ by stacking its lower diagonal elements below one another. Finally, let define the $((p+k+l) \times 1)$-vector of all the model parameters as $\boldsymbol{\Psi} = [\boldsymbol{\theta}' \quad \boldsymbol{\lambda}']'$.

2.2 *Evaluation of the likelihood*

Because of the nonlinearity of the regression function in the random effects, the likelihood of NLMEM cannot be expressed in a closed form. Indeed, for the subject i, the marginal

loglikelihood $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})$ of $\boldsymbol{\Psi}$ for the data $\boldsymbol{y_i}$ is given by:

$$\begin{aligned} L_i(\boldsymbol{y_i};\boldsymbol{\Psi}) =& log\left(\int p(\boldsymbol{y_i},\boldsymbol{\phi_i};\boldsymbol{\Psi})d\boldsymbol{\phi_i}\right) \\ =& log\left(\int p(\boldsymbol{y_i}|\boldsymbol{\phi_i};\boldsymbol{\Psi})p(\boldsymbol{\phi_i};\boldsymbol{\Psi})d\boldsymbol{\phi_i}\right) \end{aligned}$$

with $p(\boldsymbol{y_i}|\boldsymbol{\phi_i};\boldsymbol{\Psi})$ the conditional density of the observations given the random effects, $p(\boldsymbol{\phi_i};\boldsymbol{\Psi})$ the density of the individual parameters and $p(\boldsymbol{y_i},\boldsymbol{\phi_i};\boldsymbol{\Psi}))$ the likelihood of the complete data, such as:

$$\begin{aligned} -2\ log\ p(\boldsymbol{y_i},\boldsymbol{\phi_i};\boldsymbol{\Psi}) =& n_i\ log(2\pi) + log(|\boldsymbol{\Sigma_i}(\boldsymbol{X_i};\boldsymbol{\phi_i},\boldsymbol{\gamma})|) \\ &+(\boldsymbol{y_i} - f(\boldsymbol{X_i};\boldsymbol{\phi_i}))'\boldsymbol{\Sigma_i}(\boldsymbol{X_i};\boldsymbol{\phi_i},\boldsymbol{\gamma})^{-1}(\boldsymbol{y_i} - f(\boldsymbol{X_i};\boldsymbol{\phi_i})) \\ &+q\ log(2\pi) + log(|\boldsymbol{\Omega}^*|) \\ &+(\boldsymbol{\phi_i} - (\boldsymbol{\mu} + \boldsymbol{A_i\beta}))'\boldsymbol{\Omega}^{*-1}(\boldsymbol{\phi_i} - (\boldsymbol{\mu} + \boldsymbol{A_i\beta})) \end{aligned}$$

where $\boldsymbol{\Omega}^* = \boldsymbol{B}'\boldsymbol{\Omega}\boldsymbol{B}$.

Different approaches have been proposed to estimate $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})$, linearisation based methods consist of a first Taylor expansion of the model function f. The FO approach linearizes the model around the expected value of $\boldsymbol{\eta_i}$, i.e. 0. With this approximation equation (1) can be written as:

$$\boldsymbol{y_i} = f(\boldsymbol{X_i};\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}) + \left.\frac{\partial f(\boldsymbol{X_i};\boldsymbol{\phi_i})}{\partial \boldsymbol{\eta_i}}\right|'_{\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}} \boldsymbol{\eta_i} + \boldsymbol{\epsilon_i}$$

with $\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}} = \boldsymbol{\mu} + \boldsymbol{A_i\beta}$. The approximated marginal expectation $\boldsymbol{E_i}$ and variance $\boldsymbol{V_i}$ of the vector $\boldsymbol{y_i}$ are then given by:

$$\begin{aligned} \boldsymbol{E_i} &= f(\boldsymbol{X_i};\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}) \\ \boldsymbol{V_i} &= \left.\frac{\partial f(\boldsymbol{X_i};\boldsymbol{\phi_i})}{\partial \boldsymbol{\eta_i}}\right|'_{\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}} \boldsymbol{\Omega}^* \left.\frac{\partial f(\boldsymbol{X_i};\boldsymbol{\phi_i})}{\partial \boldsymbol{\eta_i}}\right|_{\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}} + \boldsymbol{\Sigma_i}(\boldsymbol{X_i};\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}},\boldsymbol{\gamma}) \end{aligned} \tag{2}$$

The loglikelihood $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})$ is approximated by:

$$-2\ L_i(\boldsymbol{y_i};\boldsymbol{\Psi}) = n_i\ log(2\pi) + log(|\boldsymbol{V_i}|) + (\boldsymbol{y_i} - \boldsymbol{E_i})'\boldsymbol{V_i}^{-1}(\boldsymbol{y_i} - \boldsymbol{E_i})'$$

Another approach is to use an importance sampling procedure (Robert and Casella, 1983)

to compute an estimate $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})_T$, of the observed loglikelihood, such as:

$$L_i(\boldsymbol{y_i};\boldsymbol{\Psi})_T = log\left(\frac{1}{T}\sum_{t=1}^{T} p(\boldsymbol{y_i}|\boldsymbol{\phi_i}^{(t)};\boldsymbol{\Psi})\frac{p(\boldsymbol{\phi_i}^{(t)};\boldsymbol{\Psi})}{h(\boldsymbol{\phi_i}^{(t)})}\right)$$

where $\boldsymbol{\phi_i}^{(t)}$ are sampled from an instrumental distribution $h(\boldsymbol{\phi_i}^{(t)})$ which is chosen in order to minimize the variance of the estimate $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})_T$. In the MONOLIX software where such a procedure is used, the instrumental function is a non centered student distribution (Samson et al., 2007; Lavielle, 2008).

2.3 *Evaluation of the estimation variance matrix*

The estimation variance matrix is composed of:

$$Var(\boldsymbol{\Psi}) = \begin{pmatrix} Var(\boldsymbol{\theta}) & Var(\boldsymbol{\theta},\boldsymbol{\lambda}) \\ Var(\boldsymbol{\theta},\boldsymbol{\lambda})' & Var(\boldsymbol{\lambda}) \end{pmatrix}$$

where $Var(\boldsymbol{\theta})$ is the $((p+k)\times(p+k))$-estimation variance matrix for the fixed effects, $Var(\boldsymbol{\lambda})$ is the $(l\times l)$-estimation variance matrix for the variance components and $Var(\boldsymbol{\theta},\boldsymbol{\lambda})$ is the $((p+k)\times l)$-estimation covariance matrix between the fixed effects and the variance components. Based on the Cramer-Rao inequality, the inverse of the Fisher estimation matrix $\boldsymbol{M_F}$ is the lower bound of the variance covariance matrix of any unbiased estimators of the parameters. In the framework of normal theory maximum likelihood, $\boldsymbol{M_F}$ is computed as the Hessian of the loglikelihood in all the model parameters and thus:

$$\boldsymbol{M_F} = \sum_{i=1}^{N} \frac{-\partial^2 L_i(\boldsymbol{y_i};\boldsymbol{\Psi})}{\partial\boldsymbol{\Psi}\partial\boldsymbol{\Psi}'}$$

where N is the total number of subjects. Usually, one assumes that the asymptotic is reached and compute $Var(\boldsymbol{\Psi})$ as equal to ${\boldsymbol{M_F}}^{-1}$.

As we need its expression in the following, let us write here the estimation variance for the fixed effects. Using the first order linearisation of the model around $\boldsymbol{\eta_i} = 0$ where $\boldsymbol{E_i}$ and $\boldsymbol{V_i}$ are given in equation (2), we have:

$$Var(\boldsymbol{\theta}) = \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}'}{\partial\boldsymbol{\theta}} {\boldsymbol{V_i}}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial\boldsymbol{\theta}}\right)^{-1} \quad (3)$$

A Taylor expansion of the model function f can also be performed around the individual parameter estimates $\hat{\boldsymbol{\phi}}_{\boldsymbol{i}} = \hat{\boldsymbol{\mu}} + \boldsymbol{A_i}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\eta}}_{\boldsymbol{i}}$, which provides a finer approximation. Then, equation (1) is rewritten as:

$$\boldsymbol{y_i} = f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}) + \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} (\boldsymbol{\phi_i} - \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}) + \boldsymbol{\epsilon_i}$$

which is reformulated for sake of simplicity in:

$$\begin{aligned}\tilde{\boldsymbol{y}}_{\boldsymbol{i}} =& \boldsymbol{y_i} - f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}) + \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} \hat{\boldsymbol{\phi}}_{\boldsymbol{i}} \\ =& \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} \boldsymbol{\phi_i} + \boldsymbol{\epsilon_i}\end{aligned}$$

Then, the approximated marginal expectation $\boldsymbol{E_i}$ and variance $\boldsymbol{V_i}$ of the vector $\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$ are given by:

$$\begin{aligned}\boldsymbol{E_i} =& \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} (\mu + \boldsymbol{A_i}\beta) \\ \boldsymbol{V_i} =& \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} \boldsymbol{\Omega}^* \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} + \boldsymbol{\Sigma_i}(\boldsymbol{X_i}, \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}, \boldsymbol{\gamma})\end{aligned}$$

With the latter approximation, $Var(\boldsymbol{\theta}, \boldsymbol{\lambda}) = 0$. This approach is implemented in the MONOLIX software.

3. Asymptotic tests and alternatives

3.1 *Asymptotic tests*

In asymptotic conditions, testing the null hypothesis H_0 : $\boldsymbol{C\theta} = 0$ on the estimates of the fixed effects $\hat{\boldsymbol{\theta}}$ can be carried out with the usual Wald chi-square test. Here, $\boldsymbol{C}$ is a $((p+k) \times U)$-contrast matrix and U the number of contrasts one wish to test. If H_0 is true, then:

$$W = (\boldsymbol{C}\hat{\boldsymbol{\theta}})'(\boldsymbol{C}'Var(\boldsymbol{\theta})\boldsymbol{C})^{-1}(\boldsymbol{C}\hat{\boldsymbol{\theta}}) \sim \chi^2_U \tag{4}$$

H_0 can also be tested using the likelihood ratio test (LRT) that compares the models with $\boldsymbol{C\theta} = 0$ or not. If H_0 is true, then:

$$S = -2 \times (L_{reduced} - L_{full}) \sim \chi^2_U$$

where $L_{reduced}$ and L_{full} are the loglikelihood of the models with $\boldsymbol{C\theta} = 0$ or not.

3.2 *Permutation based alternative*

To perform permutation tests, R data sets are generated by permuting the rows of the covariates matrix from the original data set. For a given test, one statistic Q^{obs} is estimated from the original data and one statistic Q^{perm} is estimated from each of the R data sets. Thus, we obtain r=1,...,R Q^{perm_r}, which constitute a distribution of the statistic under the null hypothesis of no covariate effect. The permutation p-value is the proportion: $(card(Q^{perm_r} \geqslant Q^{obs}) + 1)/(R + 1)$.

3.3 *F-distribution based alternative*

Comparing W/U to a F-distribution with an infinite denominator df is equivalent as using the classical Wald test. However, in biology studies, sample size is often small to moderate, thus approximate F-distributions with numerator df=U and non infinite denominator df have been proposed to correct for the departure from the asymptotic. In this section,we consider four different F-distribution based approaches.

The first approach derives from the Degrees of Freedom decomposition in the ANOVA ($=DF_{ANOVA}$) (Pinheiro and Bates, 2000) with a denominator df equal to $\sum_{i=1}^{N} n_i - (N + p + k - U)$. In the second approach, the computation of the degrees of Freedom is based on the number of RANDom effects ($=DF_{RAND}$) (Wolfinger, 2000) with a denominator df equal to $N - q$. The third approach comes from the Multivariate NonLinear Model framework ($=DF_{MNLM}$) (Gallant, 1975; Vonesh and Chinchilli, 1997) and is close to the population averaged approach which focus on the marginal expectation of the response variable. In his study on nonlinear regressions contemporaneously but not serially correlated, Gallant (1975) observed that estimation variances were underestimated and thus recommended to multiply $Var(\boldsymbol{\theta})$ by a factor $N/(N - p - k)$ and use a denominator df equal to $N - p - k$.

The fourth approach is an extension to NLMEM of the method developed by Fai and Cornelius (1996) (=DF_{FC}). Fai and Cornelius (1996) proposed to use the spectral decomposition of $\boldsymbol{C}'Var(\boldsymbol{\theta})\boldsymbol{C}$ in equation (4) to decompose W in a sum of u=1,..,U squared student statistics with df:

$$\upsilon_u = \frac{2(c_u' Var(\boldsymbol{\theta}) c_u)^2}{Var(c_u' Var(\boldsymbol{\theta}) c_u)}$$

where $Var(c_u' Var(\boldsymbol{\theta}) c_u)$ is approximated by the delta method using the estimates of $Var(\boldsymbol{\lambda})$ and the estimates of the model parameters:

$$Var(c_u' Var(\boldsymbol{\theta}) c_u) \approx \frac{\partial c_u' Var(\boldsymbol{\theta}) c_u{}'}{\partial \boldsymbol{\lambda}} Var(\boldsymbol{\lambda}) \frac{\partial c_u' Var(\boldsymbol{\theta}) c_u}{\partial \boldsymbol{\lambda}} \tag{5}$$

When $U > 1$, Fai and Cornelius (1996) found that W/U follows an approximate F-distribution with denominator df:

$$\upsilon = \frac{2\sum_{u=1}^{U} \frac{\upsilon_u}{\upsilon_u - 2}}{\sum_{u=1}^{U} \frac{\upsilon_u}{\upsilon_u - 2} - U}$$

As in NLMEM $Var(\boldsymbol{\theta})$ includes derivatives, we use the following property of matrix derivatives $A^{-1\prime} = -A^{-1}A'A^{-1}$ to derive $c_u' Var(\boldsymbol{\theta}) c_u$ in equation (5) with respect to each element of the vector $\boldsymbol{\lambda}$:

$$\frac{\partial c_u' Var(\boldsymbol{\theta}) c_u}{\partial \boldsymbol{\lambda}} = c_u' \left(-Var(\boldsymbol{\theta}) \frac{\partial \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}} \right)}{\partial \boldsymbol{\lambda}} Var(\boldsymbol{\theta}) \right) c_u \tag{6}$$

We also use this property to derive, $\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}$ in equation (6) with respect to each element of $\boldsymbol{\lambda}$:

$$\frac{\partial \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}} \right)}{\partial Vech(\boldsymbol{\Omega})} = \sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \left(-\boldsymbol{V_i}^{-1} \left. \frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}} \right|_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}}' \left. \frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}} \right|_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} \boldsymbol{V_i}^{-1} \right) \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}'$$

$$\frac{\partial \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}} \right)}{\partial a} = \sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \left(-\boldsymbol{V_i}^{-1} diag(1 + \frac{\hat{b} f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_{\boldsymbol{i}})}{\hat{a}}) \boldsymbol{V_i}^{-1} \right) \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}'$$

$$\frac{\partial \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}} \right)}{\partial b} = \sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \left(-\boldsymbol{V_i}^{-1} diag(\frac{\hat{a} f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_{\boldsymbol{i}})}{\hat{b}} + f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_{\boldsymbol{i}})^2) \boldsymbol{V_i}^{-1} \right) \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}'$$

In these derivations, we use linearisation of the model around the individual parameters estimates as described in section 2.3.

4. Real Data and Simulation Study

4.1 *Real data*

We illustrate the different approaches on data from a PK substudy of the COPHAR2-ANRS 111 study, a multicenter noncomparative pilot trial of early therapeutic drug monitoring in HIV-positive patients naïve of treatment. The objective of the trial was to assess the benefit of a pharmacological intervention after measurement of trough plasma concentrations of protease inhibitors (Duval et al., 2009). We focus on the PK sub-study from the group of patients receiving indinavir at a dose of 400, 600 and 800 mg boosted with ritonavir at a dose of 100 mg b.i.d. The coadministration of ritonavir, whose molecular structure leads to CYP3A inhibition, therefore enhances exposure to indinavir. Patients were genotyped for the exons 21 and 26 of the ABCB1 gene which code for the P-glycoprotein, as these SNPs were found to have an influence on the PK of protease inhibitors (Fellay et al., 2002; Solas et al., 2007). Polymorphisms on gene coding for proteins involved in the metabolism of indinavir were also investigated, such as the *CYP* 3A4*1B polymorphism and the *CYP* 3A5*3 and *6 polymorphisms which were found to impact on the clearance of indinavir (Anderson et al., 2006).

Pharmacokinetic profiles were obtained for 40 patients (27 men, 13 women) with an average age of 36.5 years. These profiles were determined at 1, 3, 6, and 12h after administration of the drug 2 weeks after the treatment onset.

4.2 *Simulation setting*

An extended description of the simulations can be found in Bertrand et al. (2008) and is briefly summarized below.

Parameters from the simulated NLMEM are set based on a preliminary analysis of the indinavir data without covariates. The concentrations at time t are simulated using a one compartment model at steady state ($\tau = 12\ h$) with first order absorption ($k_a = 1.4\ h^{-1}$), first order elimination ($k = 0.2\ h^{-1}$) and apparent volume of distribution ($V/F = 102\ L$).

$$f(t; k_a, k, V/F) = \frac{D}{V/F}\frac{k_a}{k_a - k}\left(\frac{e^{-kt}}{1 - e^{-k\tau}} - \frac{e^{-k_a t}}{1 - e^{-k_a \tau}}\right)$$

We use a diagonal $\mathbf{\Omega}$ matrix and a proportional error model ($g(t; \boldsymbol{\phi_i}, \boldsymbol{\gamma}) = bf(t, \boldsymbol{\phi_i})$) setting the dose (D) to 400 mg. The model parameters are expressed in term of natural logarithms to achieve the positivity requirement of pharmacokinetic parameters, $\boldsymbol{\mu} = [log(k_a) \quad log(k) \quad log(V/F)]'$ with random effect standard deviations set to 0.113, 0.41 and 0.26, respectively and the coefficient of variation for the residual error set to 20% (b=0.2).

We simulate a diplotype of SNP_1 and SNP_2 the distribution of which mimics that of exon 26 and exon 21 of the ABCB1 gene as reported by Sakaeda et al. (2002) yielding for SNP_1 unbalanced frequencies of 24%, 48% and 28%, respectively for the rare homozygotes, heterozygotes and common homozygotes. As in the intestine the P-glycoprotein restricts drug entry into the body, we consider an effect on the drug bioavailability (F) through the volume of distribution V/F. This is a simplification of the effect of P-gP on a drug pharmacokinetics as it could also alter the distribution in peripheral compartments or the elimination through entero-hepatic cycle or biliary pathway.

The genetic coefficient values are chosen to be consistent with results found in the literature for ABCB1 polymorphisms on drugs disposition (Marzolini et al., 2003) and provide clinically relevant effect, with V/F and CL/F ($=k \times V/F$) increasing from 105.4 to 200.5 L and 21.1 to 40.1 L/h respectively between common and rare homozygotes for SNP_1.

In this work, the tests under study assess only the effect of SNP_1 on V/F even if we simulated diplotypes. Thus, in the model including the covariate, the effect coefficient vector

is $\boldsymbol{\beta} = [\beta_1 \quad \beta_2]'$ with A_i equal to $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}'$ and $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}'$ for the heterozygotes and the rare homozygotes, respectively. As we test for the effect of the SNP as a whole, we have $U = 2$ and $\boldsymbol{C} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}'$.

Two hundred data sets are simulated with N=40 subjects and n=4 samples as in the COPHAR2 trial under both the null hypothesis of no genetic effect (H_0) and the alternative hypothesis of a genetic effect (H_1). Figure 1 displays spaghetti plots of simulated concentrations *versus* time for one simulated data set respectively under H_0 and under H_1, with the subjects sorted according to their genotype for SNP_1. The profile for the mean parameters, represented by a thick line, shows the relevant drop in bioavailability simulated in rare homozygotes under H_1.

[Figure 1 about here.]

5. Evaluation

5.1 *Evaluation of the simulation study*

In the present work we used the softwares NONMEM version V and MONOLIX version 2.1. The two alternatives proposed in the present work are evaluated in terms of type I error and power. The type I error estimate is the percentage of the 200 data sets simulated under H_0 where the test is significant and the power estimate is the same percentage but for the 200 data sets simulated under H_1. The prediction interval around 5% for 200 data sets is [2%; 8%]. The estimates of type I error and power obtained with permutation or F-distributions are compared to values obtained using the asymptotic tests, i.e. a theoretical threshold and using a simulation based correction, i.e. a threshold built from simulations under the model with $\boldsymbol{C\theta} = 0$.

As described in section 3, the theoretical type I error of both tests is obtained using a

threshold that is the 95^{th} percentile of a chi-square distribution with 2 degrees of freedom, i.e 5.99. In (Bertrand et al., 2008), we have evaluated the theoretical type I error of both tests on 1000 data sets simulated under H_0 using the same simulation setting as described in section 4 of the present paper. Those data sets were fitted with the models including and not including a genetic effect, thus defining the distributions of both the Wald test and the LRT statistics under the null hypothesis. The 95^{th} percentiles of those distributions are what we use in the following as thresholds by simulations, i.e 17.36 and 24.13 with FO and 7.80 and 6.96 with SAEM, for the Wald test and the LRT respectively.

If the test statistic does follow its reference distribution, the p-values of the test on 200 data sets simulated under H_0 should follow an uniform distribution on the interval $[0, 1]$. Thus, we computed the Kolmogorov distance between the uniform distribution on the interval $[0, 1]$ and the p-value distributions observed for the Wald test and the LRT under their asymptotic form and using the proposed alternatives with both estimation algorithms (Ruckdeschel et al., 2006). For a sample of 200 values, the Kolmogorov-Smirnov (K-S) test of equality of one-dimensional probability distributions is conclusive at a level of 5% when the Kolmogorov distance is below $\frac{1.358}{\sqrt{200}} = 0.096$.

In this evaluation, for the permutation test the number R of permutations is set to 1000. For the likelihood estimation in MONOLIX the number of iteration for the importance sampling T is set to a value that provides a balance between estimate accuracy and computing time (Bertrand et al., 2009b).

The fourth approach based on F-distribution (DF_{FC}) is evaluated only with SAEM in MONOLIX. We did not program the extension with the FO algorithm as we had no easy access to the NONMEM code.

5.2 *Application to real data*

The indinavir concentrations are analyzed with the same PK model as in the simulation study using the SAEM algorithm, because of its better statistical properties. The details of the covariate model building strategy are described in Bertrand et al. (2009a). Briefly, we tested for demographic and biological covariates along with the genetic polymorphisms. A first screening on the empirical Bayes estimates of the individual parameters was performed with the Wilcoxon non-parametric test for categorical variables and the Spearman non-parametric correlation test for continuous variables. Then, a forward selection based on LRT was performed on the covariates significantly associated with a parameter at the screening step. In the present work, the p-value of the covariates remaining in the final model were assessed using the asymptotic LRT, a permutation based LRT and the four F-distribution based Wald tests.

6. Results

6.1 *Simulation study*

The 200 data sets simulated under H_0 and H_1, are fitted with both FO and SAEM and the different approaches to test for a gene effect on V/F are applied. With FO, no estimates are obtained for the LRT in 4 of the 200 data sets under H_0 and H_1 respectively, because we could not obtained convergence for both models. For the Wald test, we could not obtain convergence and/or estimation variance matrix of the model including the covariate in 11 and 5 of the 200 data sets under H_0 and H_1 respectively. Besides, we could not performed the permutation test on 13 and 14 of the remaining data sets evaluated under H_0 and H_1, because convergence and/or estimation variance matrix could not be obtained in at least 950 permuted data sets.

Table 1 displays the threshold and the denominator df obtained under H_0 along with the

number of data sets on which the test could be performed and the percentage of data sets where the test is found significant under both H_0 and H_1 using the FO and SAEM algorithms with the asymptotic test, the simulation based correction, the permutation test and the four F-distribution based approaches.

[Table 1 about here.]

The thresholds used for the asymptotic test, the simulation based correction and the first three F-distribution based approaches are the same for the 200 data sets under both hypotheses. For the permutation test and the DF_{FC} approach, the threshold is computed for each data set given the distribution built from permutations or the denominator df estimated following behind the SE computation within MONOLIX code, respectively. Also the median and range of threshold and denominator df obtained under H_1 are very similar to that obtained under the null hypothesis, as expected (data not shown).

With asymptotic tests, the type I error for the Wald test and the LRT are significantly inflated with estimates about 25% and 50% with FO and 10% and 9% with SAEM. For the permutation test, the type I error estimates of the Wald test and the LRT are non-significantly different from the nominal level of 5%, with both algorithms. Yet, the range of the thresholds with FO is 10 times larger than the range obtained with SAEM for both tests. The permutation test power for the Wald test and the LRT are much lower using FO, respectively 8.3% and 51.5%, than with SAEM, respectively 73% and 72%. In a previous study, we linked this lack of power to a high correlation between the estimation errors of the parameter of the gene effect and their estimates, leading to decreased values of the Wald statistic and reducing the power to detect a genetic polymorphism effect (Bertrand et al., 2009b). This pattern is not observed using SAEM. The same decrease in power is observed with the simulation based correction for the FO algorithm. The permutation of the genotypes vector can not overcome this problem of the estimation method.

Figure 2 displays the Kolmogorov distance for the Wald test and the LRT under their asymptotic form and using the proposed alternatives with both estimation algorithms.

[Figure 2 about here.]

For the permutation test and the simulation based correction, the Kolmogorov distance estimates are below the K-S threshold for both tests and both algorithms, with the exception of the Wald test by permutations using FO. The type I error estimate of the DF_{ANOVA} approach is very close to that of the asymptotic test and their p-value distributions are significantly different from the uniform distribution with both estimation algorithms. Indeed, the large denominator df of the DF_{ANOVA} approach leads to a threshold quite close to that of a χ^2 divided by 2. The range of denominator df obtained with the DF_{FC} method is quite narrow and happens to include that of the DF_{RAND} approach. Both approaches show Kolmogorov distances above 0.096 and inflated type I error estimates though lower than the asymptotic test. With SAEM, the inflation observed for the asymptotic Wald test is corrected only with the DF_{MNLM} approach which combines a denominator df based on N and p+k plus a correction of $Var(\boldsymbol{\theta})$. The Kolmogorov distance of the DF_{MNLM} approach is similar to that obtained using permutation test or simulations based correction and they are below 0.096. With FO, the DF_{MNLM} approach does not correct for the type I error inflation of the Wald test.

In order to determine the value of denominator df that leads to exactly 5% of nominal level for the usual test (as defined in equation (4)) and the range of denominator df that lead to the predicted interval around 5%, we estimated the type I error of the Wald test on the 200 data sets simulated under H_0 trying all the integer values between 1 to 40 as denominator df with SAEM. Actually, one should use a denominator df of 10 to obtain exactly a type I error of 5%, and between 6 and 27 to obtain a type I error estimate non significantly different from 5%.

6.2 *Real data*

The indinavir concentration time data are adequately fitted by a one compartment model parameterized in first-order absorption rate, apparent volume of distribution and apparent elimination clearance (Cl/F), using SAEM. The model with no covariate has an absorption constant of 1.3 h^{-1} with an important interindividual variability of 118%, an elimination clearance of 21.9 $L.h^{-1}$ and a volume of distribution of 93.9 L with interindividual variabilities of 34.4% and 19.3%, respectively. The standard deviations of the random effect can be expressed as coefficient of variation of the pharmacokinetic parameters as in the model the latter are coded in natural logarithms. Those estimates are in accordance with the literature on indinavir given in combination with ritonavir. The coefficient of variation for the residual error is 44.5%. All of the estimation relative standard errors (RSE) were below 25% with the exception of k_a and V/F (around 30 and 60%, respectively). In this analysis, we chose not to include ritonavir as a covariate in the indinavir model, as performed in previous studies, because such parameterisation assumes a unidirectional influence of ritonavir on indinavir, which is not true (Bertrand et al., 2009a).

After the ascending selection using the asymptotic LRT, only an effect of the *CYP* 3A4*1B polymorphism effect on k_a (p-value=0.02) and an age effect on Cl/F (p-value=0.03) remain in the model. The p-values of the permutation tests are 0.04 and 0.1 for the *CYP* 3A4*1B polymorphism effect and the age effect, respectively. The corresponding p-values are 0.014 and 0.043 using the DF_{ANOVA} approach, 0.018 and 0.048 using the DF_{RAND} approach, 0.027 and 0.064 using the DF_{MNLM} approach, and 0.02 and 0.047 using the DF_{FC} approach.

Only the effect of the *CYP* 3A4*1B polymorphism remains in the final model, as the effect of age is discarded based on the p-value estimates from the permutation test and the DF_{MNLM} approach.

7. Discussion

Previous comparisons of the permutation based approach and the use of F-distribution have been performed in the linear mixed effect framework (Lin and Heagerty, 2004; Routledge, 1997). In particular Routledge (1997) concluded that when unable to guarantee, in advance, the reliability of p-values based on the F-distribution one should use a permutation test. However, to our knowledge such an evaluation has not yet been published in the specific context of NLMEM. Among the four F-distribution based approaches, that we evaluate, two are implemented in software dedicated to NLMEM, one was developed in the framework of multivariate nonlinear models and one is an extension of a method widely used in LMEM.

Simulation based correction can prove to be difficult, as for instance in studies including patients with very different dosing schemes. In such situations, permutation tests appear attractive. Their evaluation incurs a high computational burden. The decision to use the FO algorithm to represent the linearisation based methods was based on the difficulty we met in terms of achieving convergence using the FOCE-I algorithm in NONMEM. To evaluate both tests, the choice of 1000 for the number of permutations was based on the results from Manly (1998) and the decision to estimate the type I error and power on 200 data sets only provides a balance between estimation precision and reasonable computing times. The permutation test is found to be a robust alternative to correction based on simulations under H_0 from a model, with type I error close to 5%. With regard to the main assumption of permutation tests, our simulation setting ensures exchangeability of the observations as the genetic covariate was only affecting the fixed effects. The external validity of a permutation test is often questioned as it is by construction "sample dependent". Manly (1998) argues however that it is equivalent to question the representativeness of the study sample which also conditions the external validity of classical tests.

Because the covariates enter the model in a linear fashion, we also consider the DF_{FC},

a widely used method in LMEM. This approach is sometime criticized as the validity of the chi-square (and by extension the Fisher) approximation to the statistic distribution depends on not testing at the boundary and on having a large sample size. Thus, rather than approximating a difficult distribution by a more common distribution one might prefer to use parametric bootstrap to get a reference distribution. In favor of using F-distribution is the fact that the shapes of the 200 distributions obtained by permutation for the Wald test with SAEM are quite close to a Fisher distribution (results not shown). The DF_{FC} method however fails to correct for the Wald type I error inflation, as the denominator df estimates for the 200 data sets are too large. One limit of the present work is that we consider the method developed by Fai and Cornelius (1996) and not the method developed by Kenward and Roger (1997). We do not consider the latter because it is based on restricted maximum likelihood (REML) which is not commonly used in NLMEM. In both NONMEM and the NLMIXED Macro in SAS, only maximum likelihood estimation is implemented. REML estimation for NLMEM can be performed using the nlme function in R and Meza et al. (2007) have developed a REML extension of the SAEM algorithm not implemented in MONOLIX yet. A perspective of this work would be to assess the Kenward and Roger (1997) method using the REML extension of SAEM proposed by Meza et al. (2007). As for the present work, the Fai and Cornelius (1996) method's has been shown to perform similarly to the Kenward and Roger (1997) method's when a simple variance-covariance matrix is used as in our study (Schaalje et al., 2002). On this example, the only method based on F-distribution that corrects for the type I error inflation of the Wald test is the DF_{MNLM} method. This method was recommended by Gallant (1975) based on a Monte Carlo study in the multivariate nonlinear model framework. In the previous study evaluating the Wald and the LRT and the simulation-based corrections, we considered two other designs (Bertrand et al., 2009b): a design optimized using the PFIM software (Retout et al., 2007) including

N=80 subjects sorted in 4 groups with n=2 samples and a combined design with N=20 subjects having n=4 samples plus N=80 subjects with only a trough concentration. The type I error estimates using the classical Wald test and the DF_{MNLM} method are 8.7% and 6.1%, respectively for the design N=80/n=2 and 8.4% and 5.7%, respectively for the design N=100/n=4,1. These results show that the DF_{MNLM} method also corrects the type I error inflation observed in these additional designs. It should however be investigated in designs involving different number of samples per subject and in other simulation settings.

With the FO algorithm, the DF_{MNLM} method does not correct the type I error inflation of the Wald test. Moreover, the FO algorithm shows an important lack of power using permutation test and simulation based correction, especially for the Wald test. The departure from the asymptotic with the FO method appears to be too strong for the DF_{MNLM} method to correct the inflation and for the permutation test to insure a high power. As for the conclusive K-S test for the Wald test by permutations with FO, we hypothesize this may derive from the multiplicity of the tests.

A limitation of this work is that we did not use the AGQ algorithm. Still, as this algorithm is an exact method we assume that the results would have been similar to those obtain with the SAEM algorithm.

In the application we show an effect of the *CYP* 3A4*1B on the absorption of indinavir, the age effect on clearance being withdrawn from the model after the backward selection based on the permutation test or the DF_{MNLM} approach (Bertrand et al., 2009a). We hypothesize that in *CYP* 3A4*1B*1B patients, the ritonavir inhibition potency is lowered, leading to a higher first pass effect of indinavir, although this does not impact its clearance.

This work confirms the feasibility of permutation tests in pharmacogenetic studies and invites to further investigate the F-distribution based approach recommended in multivariate nonlinear model in other simulation settings.

Acknowledgements

The authors would like to thank the IFR02 of INSERM and Hervé Le Nagard for the use of the "centre de biomodélisation" as well as the Pr. Marc Lavielle for the precious help he provided in using MONOLIX.

During this work, Julie Bertrand was supported by a grant from the Institut de Recherches Internationales Servier (France).

References

Anderson, P., Lamba, J., Aquilante, C., Schuetz, E., and Fletcher, C. (2006). Pharmacogenetic characteristics of indinavir, zidovudine, and lamivudine therapy in HIV-infected adults: a pilot study. *Journal of Acquired Immune Deficiency Syndromes* **42,** 441–49.

Bertrand, J., Comets, E., Laffont, C., Chenel, M., and Mentré, F. (2009b). Pharmacogenetics and population pharmacokinetics: impact of the design on three tests using the SAEM algorithm. *Journal of Pharmacokinetics and Pharmacodynamics* **36,** 317–339.

Bertrand, J., Comets, E., and Mentré, F. (2008). Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *Journal of Biopharmaceutical Statistics* **18,** 1084–1102.

Bertrand, J., Tréluyer, J., Panhard, X., Tran, A., S, Rey, E., Salmon-Céron, D., Duval, X., Mentré, F., and the COPHAR2-ANRS 111 study group. (2009a). Influence of pharmacogenetics on indinavir disposition and short-term response in HIV patients initiating HAART. *European Journal of Clinical Pharmacology* **65,** 667–678.

Deylon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of EM algorithm. *Annals of Statistics* **27,** 94–128.

Ding, A. A. and Wu, H. (2001). Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics* **2,** 13–29.

Duval, X., Mentré, F., Rey, E., Auleley, S., Peytavin, G., Biour, M., Métro, A., Goujard, C., Taburet, A., Lascoux, C., Panhard, X., Tréluyer, J., and Salmon, D. (2009). Benefit of therapeutic drug monitoring of protease inhibitors in HIV-infected patients depends on PI used in HAART regimen - ANRS 111 trial. *Fundamental & Clinical Pharmacology* **23,** 491–500.

Fai, A. H. T. and Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computing and Simulation* **54,** 363–378.

Fellay, J., Marzolini, C., Meaden, E., Back, D., Buclin, T., and Chave, J. (2002). Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter gene MDR1: a pharmacogenetic study. *Lancet* **359,** 30–36.

Gabrielsson, J. and Weiner, D. (1999). *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*. Stockholm: Apotekarsocieteten.

Gallant, A. R. (1975). Seemingly unrelated nonlinear regressions. *Journal of econometrics* **3,** 35–50.

Girard, P. and Mentré, F. (2005). A comparison of estimation methods in nonlinear mixed effects models using a blind analysis. 14^{th} Population Approach Group in Europe.

Good, P. (1994). *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer, New York.

Guedj, J., Thiébaut, R., and Commenges, D. (2007). Maximum likelihood estimation in dynamical models of HIV. *Biometrics* **63,** 1198–1206.

Holford, N. (2001). Randomisation test methods. *A NONMEM user's perspective* .

Jacqmin-Gadda, H., Proust-Lima, C., Taylor, J. M., and Commenges, D. (2009). Score Test for Conditional Independence Between Longitudinal Outcome and Time to Event Given

the Classes in the Joint Latent Class Model. *Biometrics* .

Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53,** 983–997.

Lavielle, M. (2008). *MONOLIX (MOdèles NOn LInéaires à effets miXtes).* MONOLIX group, Orsay, France.

Lin, D. Y. and Heagerty, P. J., editors (2004). *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*. Springer, New York.

Lindstrom, M. and Bates, D. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46,** 673–687.

Manly, B. (1998). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Texts in statistical science. Chapman & Hall, London, 2 edition.

Marzolini, C., Paus, E., Buclin, T., and Kim, R. B. (2003). Polymorphisms in human MDR1 (p-glycoprotein): recent advances and clinical relevance. *Clinical Pharmacology and Therapeutics* **75,** 13–33.

Meza, C., Jaffrézic, F., and Foulley, J. L. (2007). REML estimation of variance parameters in nonlinear mixed effects models using the SAEM algorithm. *Biometrical journal* **49,** 876–888.

Pillai, G. C., Mentré, F., and Steimer, J. L. (2005). Non-linear mixed effects modeling - from methodology and software development to driving implementation in drug development science. *Journal of Pharmacokinetics and Pharmacodynamics* **32,** 161–183.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed Effects Models in S and S-Plus*. Springer Verlag, New York.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Retout, S., Comets, E., Nagard, H. L., Bazzoli, C., and Mentré, F. (2007). *PFIM Interface*

2.1. UMR738, INSERM, Université Paris 7, Paris, France.

Robert, C. P. and Casella, G. (1983). *Monte Carlo Statistical Methods.* New York: Springer Verlag.

Routledge, R. D. (1997). P-values from permutation and F-tests. *Computational Statistics & Data Analysis* **24,** 379–386.

Ruckdeschel, P., Kohl, M., Stabla, T., and Camphausen, F. (2006). S4 classes for distributions. *R News* **6,** 2–6.

Sakaeda, T., Nakamura, T., and Okumura, K. (2002). MDR1 genotype-related pharmacokinetics and pharmacodynamics. *Biological & Pharmaceutical Bulletin* **25,** 1391–1400.

Samson, A., Lavielle, M., and Mentré, F. (2007). The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Statistics in Medicine* **26,** 4860–4875.

SAS Institute Inc. (2004). *SAS 9.1.2.* Cary, NC, USA.

Schaalje, G. B., McBride, J. B., and Fellingham, G. W. (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics* **7,** 512–524.

Sheiner, L. and Beal, S. (1998). *NONMEM Version 5.1.* University of California, NONMEM Project Group, San Francisco.

Sheiner, L., Rosenberg, B., and Melmon, K. (1972). Modelling of individual pharmacokinetics for computer aided drug dosage. *Computers and Biomedical Research* **5,** 441–459.

Solas, C., Simon, N., Drogoul, N., Quaranta, S., Frixon-Marin, V., Brunet, V. B.-R. C., JA, J. G., Durand, A., Lacarelle, B., and Poizot-Martin, I. (2007). Minimal effect of MDR1 and CYP3A5 genetic polymorphisms on the pharmacokinetics of indinavir in HIV-infected patients. *British Journal of Clinical Pharmacology* **64,** 353–362.

Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis*

of Repeated Measurements. Marcel Dekker, New York.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80,** 791–795.

Wolfinger, R. D. (2000). Fitting nonlinear mixed models with the new nlmixed procedure. Technical report, SAS Institute Inc., Cary, NC. Paper 287.

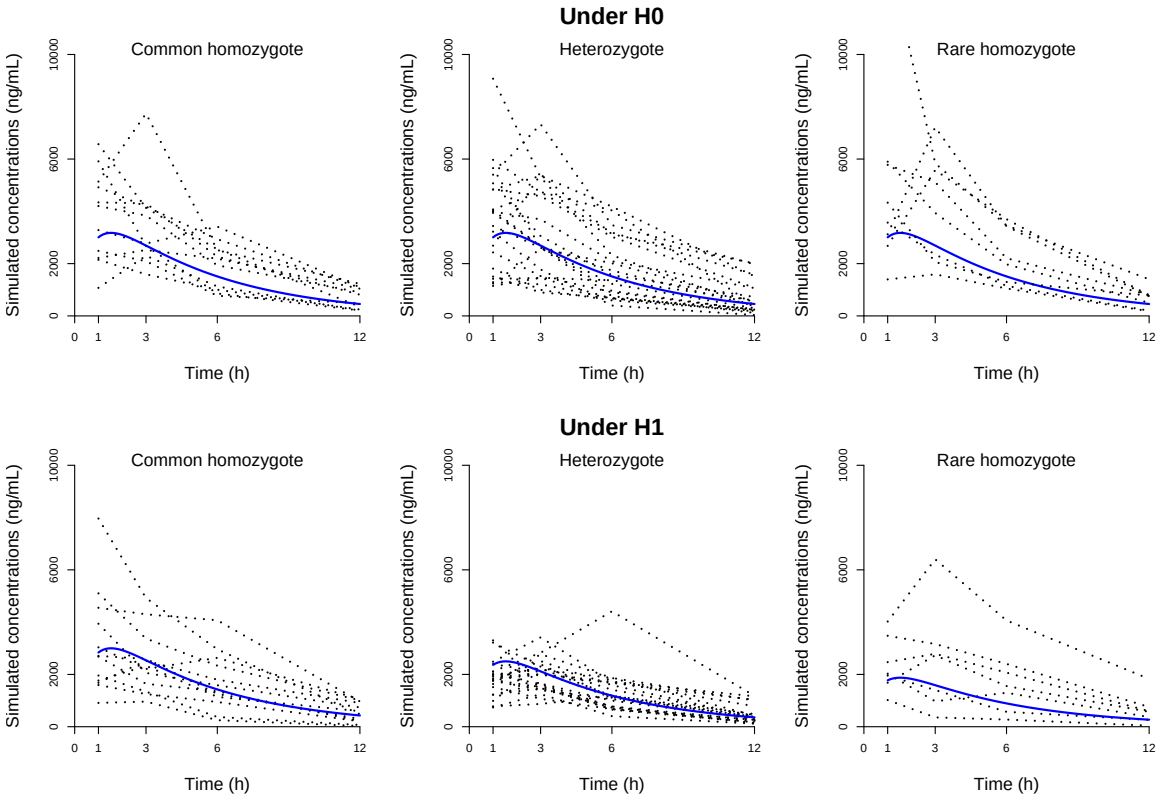


Figure 1. Concentrations (ng/mL) simulated for a representative data set under H_0 (top) and a representative one under H_1 (bottom). The subjects are sorted by genotype for the exon SNP_1: common homozygotes (left), heterozygotes (center) and rare homozygotes (right). The profile for the mean parameters is represented by a thick line. Individual concentration curves are represented by dotted lines.

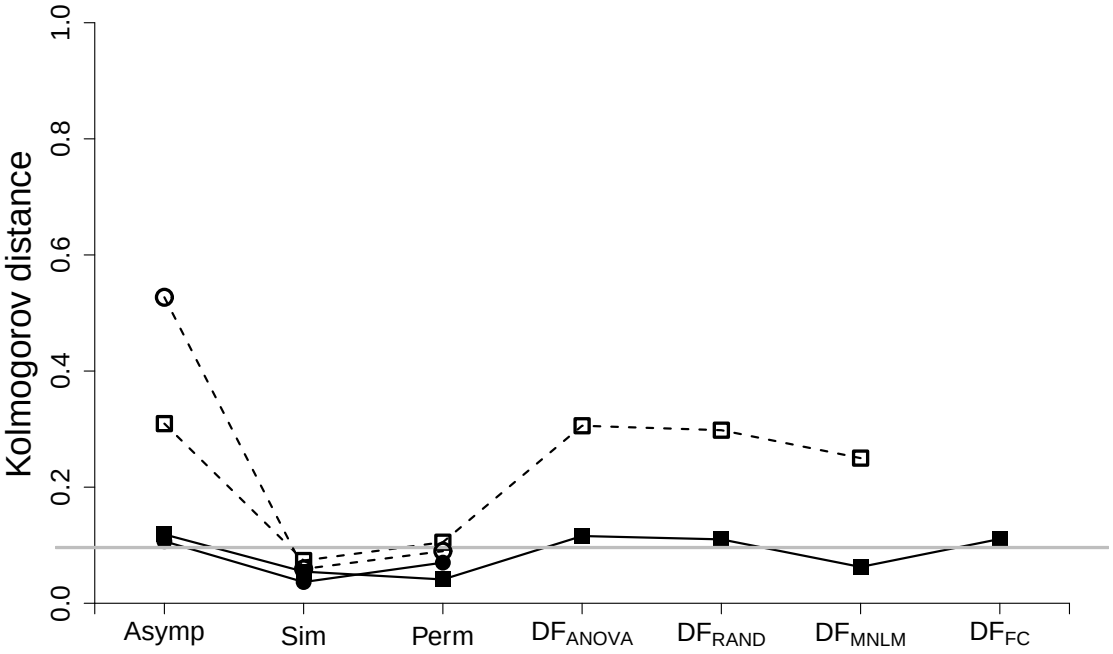


Figure 2. Kolmogorov distance between the uniform distribution on the interval $[0, 1]$ and the p-value distributions obtained on 200 simulations under H_0 for the Wald test (squares) and the LRT (circles) under their asymptotic form and using the proposed alternatives with FO (empty symbols connected by dashed lines) and SAEM (filled symbols connected by straight lines). The horizontal gray line is the corresponding threshold of the Kolmogorov-Smirnov test for 200 data sets.

Table 1

Threshold, denominator degrees of freedom (df), type I error and power estimates (%) of the Wald test and the LRT for different approaches, using both FO and SAEM

Algorithm	Test	Approach	denominator df	Threshold	T	Type I error	T	Power
FO	Wald	asymptotic	-	5.99*	189	24.3	195	62.0
		simulation	-	17.36*	189	5.8	195	9.3
		permutation	-	15.67[9.82;38.87]*†	176	5.6	181	8.3
		DF_{ANOVA}	117	3.07	189	24.3	195	59.3
		DF_{RAND}	37	3.25	189	23.3	195	57.9
		DF_{MNLM}	35	4.12	189	19.0	195	49.7
	LRT	asymptotic	-	5.99*	196	48.9	196	91.3
		simulation	-	24.13*	196	5.6	196	46.9
		permutation	-	19.32[9.97;50.82]*†	196	4.6	196	51.5
SAEM	Wald	asymptotic	-	5.99*	200	10.0	200	82.0
		simulation	-	7.80*	200	5.5	200	73.0
		permutation	-	7.33[6.33;8.46]*†	200	6.0	200	73.0
		DF_{ANOVA}	117	3.07	200	9.0	200	81.5
		DF_{RAND}	37	3.25	200	8.5	200	80.5
		DF_{MNLM}	35	4.12	200	6.5	200	73.0
		DF_{FC}	39.8[36.3;43.8]†	3.23[3.21;3.26]†	200	8.5	200	77.0
	LRT	asymptotic	-	5.99*	200	9.0	200	77.0
		simulation	-	6.96*	200	6.5	200	72.0
		permutation	-	6.73[4.45-8.11]*†	200	7.0	200	72.0

T = number of data sets on which the test could be performed.

Predicted interval for a nominal level of 5% = $[2.0 - 8.0]$.

* Threshold for the distribution of W, not W/2.

† The threshold for the permutation test and both the threshold and the denominator df for the DF_{FC} approach are displayed as median [range] over the T data sets simulated under H_0.

Chapitre 3

Travaux appliqués

3.1 Influence de la génétique sur la pharmacocinétique et la réponse à court terme de l'indinavir chez des patients naïfs d'inhibiteur de protéase

3.1.1 Résumé

Les données qui ont inspiré l'étude de simulation présentée dans les différents travaux méthodologiques au chapitre 2 proviennent de la sous-étude pharmacocinétique sur l'indinavir réalisée dans l'essai COPHAR 2-ANRS 111.

Les concentrations plasmatiques d'indinavir et de ritonavir de 42 patients naïfs de traitement par inhibiteur de protéase ont été recueillies à 1h, 3h, 6h et 12h après l'administration d'indinavir, deux semaines après le début du traitement. La décroissance de la log-charge virale était mesurée à l'entrée dans l'étude puis après 2 semaines de traitement alors que les indicateurs de toxicité métabolique étaient mesurés 4 semaines avant puis après le début du traitement. Chez chaque patient, différentes covariables démographiques étaient documentées et un génotypage était réalisé pour les exons 26 et 21 du gène ABCB1 et pour les polymorphismes *CYP* 3A5*3, *6 et *CYP* 3A4*1B.

Le ritonavir est un moindre inhibiteur de la protéase inverse mais une inhibiteur puissant des enzymes CYP 3A4 et CYP 3A5, et permet ainsi d'augmenter les concentrations d'indinavir. Nous avons modélisé séparément les données d'indinavir et de ritonavir sans inclure dans le modèle d'indinavir l'exposition au ritonavir comme cela à pu être réalisé dans d'autre études (Csajka et al., 2004; Brendel et al., 2005; Kappelhoff et al., 2005a) car cette paramétrisation suppose une action unidirectionnelle du ritonavir sur l'indinavir. Or, Kappelhoff et al. (2005b) ont montré que les concentrations de ritonavir sont plus basses lorsqu'il est administré en combinaison avec du lopinavir plutôt qu'avec de l'indinavir. Nous avons alors exploré les covariables expliquant la forte variabilité inter-sujets observée sur les paramètres pharmacocinétiques de l'indinavir et du ritonavir. Nous avons aussi étudié les relations entre la décroissance de la log-charge virale à deux semaines et les concentrations d'indinavir ainsi que la toxicité à court terme et les

concentrations d'indinavir et de ritonavir. Nos précédents travaux ayant mis en évidence une inflation de l'erreur de type I du LRT sur ce plan d'expérience, nous avons utilisé l'approche par permutation pour déterminer les p-values des covariables inclues dans le modèle final.

Les concentrations d'indinavir et de ritonavir ont été décrites par deux modèles à un compartiment d'absorption et d'élimination d'ordre 1 dont la bonne adéquation aux données a été évaluée. Nous avons mis en évidence un effet du polymorphisme du *CYP* 3A4*1B sur la constante d'absorption de l'indinavir et l'augmentation à court terme des triglycérides, possiblement à travers une diminution de la capacité inhibitrice du ritonavir chez les homozygotes *1B. Par ailleurs, la clairance du ritonavir était significativement diminuée chez les patients avec un taux plasmatique d'orosomucoïde élevé. Enfin, le caractère prédictif des concentrations d'indinavir sur l'efficacité à court terme a été confirmé.

Cette analyse a fait l'objet d'un article publié dans la revue *European Journal of Clinical Pharmacology*.

3.1.2 Article 4 (publié)

Eur J Clin Pharmacol (2009) 65:667–678
DOI 10.1007/s00228-009-0660-5

SPECIAL ARTICLE

Influence of pharmacogenetics on indinavir disposition and short-term response in HIV patients initiating HAART

Julie Bertrand · Jean-Marc Treluyer · Xavière Panhard · Agnes Tran · Solange Auleley · Elisabeth Rey · Dominique Salmon-Céron · Xavier Duval · France Mentré · the COPHAR2-ANRS 111 Study Group

Received: 12 January 2009 / Accepted: 7 April 2009 / Published online: 14 May 2009

Abstract
Aims To assess the relationship between genetic polymorphisms and indinavir pharmacokinetic variability and to study the link between concentrations and short-term response or metabolic safety.
Methods Forty protease inhibitor-naive patients initiating highly active antiretroviral therapy (HAART) including indinavir/ritonavir and enrolled in the COPHAR 2–ANRS 111 trial were studied. At week 2, four blood samples were taken before and up to 6 h following drug intake. A population pharmacokinetic analysis was performed using the stochastic approximation expectation maximization (SAEM) algorithm implemented in MONOLIX software. The area under the concentration–time curve (AUC) and maximum (C_{max}) and trough concentrations (C_{trough}) of indinavir were derived from the population model and tested for their correlation with short-term viral response and safety measurements, while for ritonavir, these same three parameters were tested for their correlation with short-term biochemical safety
Results A one-compartment model with first-order absorption and elimination best described both indinavir and ritonavir concentrations. For indinavir, the estimated clearance and volume of distribution were 22.2 L/h and 97.3 L, respectively. The eight patients with the *1B/*1B genotype for the *CYP3A*4 gene showed a 70% decrease in absorption compared to those with the *1A/*1B or *1A/*1A genotypes (0.5 vs. 2.1, P=0.04, likelihood ratio test by permutation). The indinavir AUC and C_{trough} were positively correlated with the decrease in human immunodeficiency virus RNA between week 0 and week 2 (r= 0.4, P=0.03 and r =-0.4, P=0.03, respectively). Patients with the *1B/*1B genotype also had a significantly lower indinavir C_{max} (median 3.6, range 2.1–5.2 ng/mL) than those with the *1A/*1B or *1A/*1A genotypes (median 4.4, range 2.2–8.3 ng/mL) (P=0.04)

J. Bertrand (✉)
UMR 738,
INSERM, Université Paris Diderot,
UFR de Médecine, 16, rue Henri Huchard,
75018 Paris, France
e-mail: julie.bertrand@inserm.fr

J.-M. Treluyer
EA3620,
Université Paris Descartes,
Paris, France

J.-M. Treluyer
Service de Pharmacologie Clinique,
Hôpital Cochin-Saint-Vincent-de-Paul, AP-HP,
Paris, France

X. Panhard · F. Mentré
UMR 738, AP-HP,
INSERM, Université Paris Diderot,
Paris, France

A. Tran · E. Rey
Service de Pharmacologie Clinique,
Hôpital Cochin-Saint-Vincent-de-Paul, AP-HP,
Paris, France

S. Auleley · X. Duval
UMR 738, INSERM, Université Paris Diderot,
Paris, France

D. Salmon-Céron
Service de Médecine Interne,
Hôpital Cochin-Saint-Vincent-de-Paul, AP-HP,
Paris, France

X. Panhard · F. Mentré
UF de Biostatistiques, Hôpital Bichat,
Paris, France

X. Duval
Centre d'Investigation Clinique, Hôpital Bichat, AP-HP,
Paris, France

and a lower increase in triglycerides during the first 4 weeks of treatment (median 0.1, range −0.7 to 1.4 vs. median 0.6, range −0.5 to 1.7 mmol/L, respectively; P=0.02). For ritonavir, the estimated clearance and volume of distribution were 8.3 L/h and 60.7 L, respectively, and concentrations were not found to be correlated to biochemical safety. Indinavir and ritonavir absorption rate constants were found to be correlated, as well as their apparent volumes of distribution and clearances, indicating correlated bioavailability of the two drugs.

Conclusion The *CYP3A4**1B polymorphism was found to influence the pharmacokinetics of indinavir and, to some extent, the biochemical safety of indinavir.

Keywords CYP3A4 · Efficacy · Nonlinear mixed effects modeling · Pharmacokinetics · Protease inhibitors · Safety

Introduction

Indinavir has been one of the preferred protease inhibitor (PI) included in highly active antiretroviral therapy (HAART). Even if not recommended as initial therapy, indinavir is currently still used in patients who initiated their therapy with this PI and have kept a viral load below the limit of quantification with an acceptable safety profile. Compared to others PI, indinavir exhibits a high penetration into viral reservoirs, such as genital compartments and the central nervous system (CNS) [1], and it has been determined that the better distribution of indinavir leads to better outcomes in neurological complications related to human immunodeficiency virus (HIV) [2]. The pharmacokinetics (PK) of indinavir is characterised by high maximal concentrations, leading to potential toxicity, notably nephrolithiasis [3], and low minimum concentrations with respect to the 95% inhibitory concentration of the virus. These low residual concentrations result from an extended oxidative metabolism by the cytochrome P450 (CYP) 3A isoenzyme [4]. The co-administration of ritonavir, whose molecular structure leads to CYP3A inhibition, therefore enhances exposure to indinavir [5, 6]. Ritonavir is given at a lower dose as a booster than for therapeutic use, but it has been shown nevertheless to influence metabolic profiles, especially those associated with lipid disorders [7, 8].

The large inter-patient and intra-patient variability of indinavir pharmacokinetics is well referenced [9–11]. Genetic polymorphisms partly explain this variability, as far as the proteins involved in the metabolism and transport of PI are concerned. However, few studies have investigated the impact of ABCB1 polymorphisms, a gene coding for P-glycoprotein, and *CYP3A5* and *CYP3A4**1B polymorphisms on indinavir pharmacokinetics. Solas et al. [12] reported that the ABCB1 C3435T genotype affects the absorption constant of indinavir, whereas Verstuyft et al. [13] found an absence of association. Anderson et al. [14] observed that *CYP3A5* expressors (*CYP3A5**1 carriers) have a significantly faster oral clearance than non-expressors. To date, no relationship has been found between the *CYP3A4**1B polymorphism and alterations in *CYP3A* substrate metabolism, but clinical data have shown an association between the *CYP3A**1B polymorphism and disease risk/treatment toxicity [15].

Efficacy [16, 17] as well as adverse events [3, 18] have been related to indinavir plasma concentrations. Thus, therapeutic drug-monitoring appears to be a potent tool to achieve undetectable HIV-RNA and prevent toxicity for this drug. The COPHAR 2–ANRS 111 trial is a multi-centre, non-comparative pilot trial of early therapeutic drug-monitoring in HIV-positive patients naive for PI-containing HAART [19]. We focused on the PK sub-study from the group of patients receiving indinavir boosted with ritonavir. The aims of this paper were to estimate the population PK parameters and variability of indinavir and ritonavir in HIV patients, to evaluate the impact of genetic polymorphisms on indinavir PK and to study the link between indinavir concentrations and short-term efficacy and metabolic safety.

Methods

Study

The COPHAR 2–ANRS 111 study is a multi-centre non-comparative prospective pilot trial of early-dose adaptation in HIV-positive PI-naive patients starting a PI-containing HAART treatment. The trial started on July 2002 and was completed by the end of March 2005. The objective was to assess the benefit of pharmacological advice based on trough plasma concentrations of PI. The study involved three groups treated with indinavir, nelfinavir or lopinavir, respectively. In the study reported here we analysed the data obtained during the first month of treatment in the indinavir group. A similar analysis of data in the nelfinavir group was performed by Hirt et al. [19, 20; see these papers for details].

Patients were required to have a baseline plasma viral load value >1000 copies/mL and to be PI treatment-naive. Patients were started on a HAART treatment containing 400, 600 or 800 mg of indinavir twice daily (b.i.d.) associated with ritonavir booster (100 mg b.i.d.) and two nucleoside analogues. The first dose was left to the treating physicians' discretion, and no dose adaptation was performed from week 0 (W0) to W4. A detailed PK study was performed at W2. Adherence was evaluated at W2 by means of a validated auto-questionnaire [21], and patients

were classified as adherent when they reported no shift in their treatment schedule during the last 4 days; in all other cases, they were classified as non-adherent.

Data on viral load and CD4 count were collected at baseline (D0) and at W2. Biochemical profiles of total cholesterol, high-density lipoprotein cholesterol, triglyceride and glycaemia as well as creatinine clearance and clinical events (diarrhoea grade of 2) were determined 4 weeks before treatment initiation (W−4) and at W4.

The study was performed in accordance with the Declaration of Helsinki and its amendments. All subjects provided written informed consent, and the protocol as well as the amendment for the pharmacogenetic study was approved by the Ethics Committee of the Bicêtre Hospital (France).

Indinavir and ritonavir concentration measurements

During a visit to the hospital at W2, the patients were sampled on arrival to measure trough concentrations. Patients were asked to record the time at which the dose was taken on the previous evening, given their medications, and then sampled again 1, 3 and 6 h after drug administration. Plasma concentrations were assumed to be at steady state with trough concentrations considered as following the drug intake using the delay reported by the patient the from previous dosing. Plasma concentrations were determined in the laboratories of the hospitals by a specific high-performance liquid chromatography protocol. The participant laboratories were cross-validated before starting the study. Results of the blind inter-laboratory quality control at three concentrations for indinavir and for ritonavir were within 15% of the target values for medium and high values and within 20% for low values. Lower limits of quantification (LOQ) were 0.02 mg/L for indinavir and 0.025 mg/L for ritonavir.

Genetic polymorphisms

All of the genotyping analyses were performed in the same laboratory. Total DNA was extracted from plasma samples using the QIAamp DNA Blood Mini kit (Qiagen, Courtaboeuf, France). ABCB1 polymorphisms in exons 21 (GG, GT, TT) and 26 (CC, CT, TT) were determined using previously published methods [22]. The genotyping of *CYP3A5* (*1*1, *1*3, *3*3, *1*6, *6*6) was performed by real-time PCR applying TaqMan MGB probe technology (Applied Biosystems, Foster City, CA). Genotyping for *CYP3A4* (*1B*1B, *1B*1A, *1A*1A) was determined by PCR, followed by direct sequencing. The PCR analysis was performed using a GenAmp PCR System 9700 (Applied Biosystems) according to a previously published method [23]. Amplified DNA was purified using the QiaQuick DNA Purification System (Qiagen) and sequenced using BigDye Terminator chemistry and an ABI PRISM 3100 genetic analyser (Applied Biosystems). At least two positive controls were used for each genotyping analysis: one homozygous for the wild-type allele and one heterozygous (and, when available, one homozygous) for the mutated allele. These controls were DNA that had already been sequenced.

Allele frequencies (p for the wild allele and $q=1-p$ for the mutant allele) were estimated by gene counting. Departure from Hardy–Weinberg proportions (p^2, $2pq$, q^2) was tested by a χ^2 test with 1 *df* within each ethnic group [24]. We used two approaches to define patients belonging to an ethnic group: (1) classification of the patient according to town, birth area and nationality; (2) classification by means of genotype information using the Structure software [25]. This software is based on a Bayesian approach and computes the a posteriori probabilities of each individual of belonging to a given ethnic group. We assumed each locus to be at the Hardy–Weinberg equilibrium and patients to originate in one ethnic group (with its own characteristic set of allele frequencies).

Population PK analysis

We used a population approach to analyse the concentration–time data at W2 for indinavir and for ritonavir separately. Model fitting and estimation of the population model parameters were performed using the stochastic approximation expectation maximization algorithm (SAEM) for nonlinear mixed-effects models implemented in the MONOLIX software ver. 2.1 [26–28]. Both indinavir and ritonavir concentrations were fitted by a one-compartment model with first-order absorption and first-order elimination parameterised in the absorption rate constant (k_a), oral clearance (Cl/F) and oral volume of distribution (V/F). Each model was assumed at steady state with trough concentrations considered as following the drug intake.

An exponential model was used for inter-individual variability where random effects were assumed to follow a normal distribution with zero mean and diagonal variance matrix. Additive, proportional and combined error models were tested, and model choice was based on the likelihood ratio test (LRT) and goodness-of-fit plots (observed vs. predicted population and individual concentrations; population and individual weighted residuals vs. predicted concentrations and vs. time). We performed a visual predictive check (VPC) with 1000 simulated data sets to evaluate the basic model [29].

Interaction between ritonavir and indinavir PK was evaluated with the individual parameters estimated from the basic model for each drug. All of the different correlations were tested with the Spearman non-parametric correlation test.

Assessment of the effect of covariates

The effects of the following covariates were evaluated from the basic model: dose, concomitant use of the zidovudine lamivudine combination (AZT/3TC), co-infection by hepatitis C or B (VHC/VHB), adherence as previously defined, sex, ethnic group, the four studied genetic polymorphisms (ABCB1 exon 26, ABCB1 exon 21, *CYP3A5* and *CYP3A4*) and the CDC classification for HIV infection as categorical variables; age, body mass index (BMI), body weight, creatinine clearance, albumin and orosomucoid levels as continuous variables. The latter were centered to the median and log-transformed for model interpretation convenience.

Each of the four genetic polymorphisms was analysed by means of two binary categorisations: first, wild homozygotes versus heterozygotes or mutant homozygotes; second, heterozygotes or wild homozygotes versus mutant homozygotes. Categorisation in three classes was also tested: wild homozygotes versus heterozygotes versus mutant homozygotes. Missing continuous covariates were replaced with the median, and patients with missing discrete covariates were discarded for the corresponding analysis. The effects of covariates on the empirical Bayes estimates (EBE) of each individual PK parameter from the basic model were tested with the Wilcoxon non-parametric test for categorical variables and the Spearman non-parametric correlation test for continuous variables. The population covariate model was built with the covariates, which were found to have an effect in this first step with a P value<0.1. When a genetic covariate was found to have an effect whatever the categorisation, the same categorisation as other genetic covariates also found to have an effect was chosen in model selection for consistency.

A forward selection of these covariates for the population model was performed using the LRT with a significance threshold at $P<0.05$. From this ascending method, a backward elimination procedure was performed. In order to correct the inflation of the LRT type I error on small sample size [30], the backward selection was realized using permutation [31]. More specifically, 1000 data sets are generated by permuting the rows of the covariates matrix from the original data set. For each covariate, one likelihood ratio statistic, LRT^{obs}, is estimated from the original data and one likelihood ratio statistic, LRT^{perm}, is estimated from each of the 1000 data sets. Thus, we obtain $j = 1, ..., 1000\, LRT^{perm_j}$. The permutation P value is the proportion : $card\left(LRT^{perm_j} > LRT^{obs}\right)/1000$.

Short-term efficacy and safety and link with concentrations

As there was no change of dose before W4, we studied the link between concentration at W2 and efficacy or safety during the first 2 or 4 weeks of treatment. For short-term efficacy, the difference of log viral load between the day of treatment initiation and W2 (Δ logVL) was studied. The significance of the viral load decrease was tested by a Wilcoxon non-parametric paired test.

Individual area under the concentration–time curve (AUC), maximal plasma concentration (C_{max}) and trough concentrations (C_{trough}) of indinavir at steady-state were derived for each patient using the EBE of the individual parameters from the basic model and their corresponding dose of indinavir. The relationship between indinavir dose, indinavir AUC, C_{max}, C_{trough} and Δ logVL was evaluated using the Spearman correlation test. A Wilcoxon non-parametric test was performed to compare the Δ logVL between patients with or without a C_{trough} below the lower limit of the therapeutic range used in the COPHAR 2–ANRS 111 trial: 150 ng/mL.

Safety was analysed by determining the difference between 4 weeks before and 4 weeks after treatment initiation in terms of total cholesterol (ΔTC), high-density lipoprotein cholesterol (ΔHDL), triglyceride (Δtrig) and glycaemia (Δgly) and also by the appearance of diarrhoea (grade 2) between treatment initiation and W4. To the best of our knowledge, no precocious biological markers exist for nephrolithiasis; however, creatinine clearance has been found to relate to the occurrence of severe adverse events (including nephrolithiasis) in a multivariate analysis [3]. Thus, we also analysed the difference in creatinine clearance (ΔClCr), computed with the Cockcroft–Gault formula using body weight and serum creatinine 4 weeks before and 4 weeks after treatment initiation. The significance of these differences was tested using a Wilcoxon non-parametric paired test.

We performed Spearman correlation tests between indinavir dose, indinavir AUC, C_{max} , C_{trough} and ΔTC, ΔHDL, Δtrig, Δgly and ΔClCr. We used Wilcoxon non-parametric tests to compare these differences between patients with or without an indinavir C_{trough} over the upper limit defined in the therapeutic index (550 ng/mL). We studied the link between the appearance of grade 2 diarrhoea (yes/no) between treatment initiation and W4 and indinavir dose, indinavir AUC, C_{max} and C_{trough} using a Wilcoxon non-parametric test, and we studied the association with or without an indinavir $C_{trough} > 550$ ng/mL using a Fisher exact test.

We assessed the relation between the genetic polymorphisms remaining in the final population model and indinavir dose, indinavir AUC, C_{max} and C_{trough} and the relation between these genetic polymorphisms and the short-term efficacy and safety outcomes using Wilcoxon non-parametric tests.

We also derived AUC, C_{max} and C_{trough} for ritonavir and performed Spearman correlation tests with ΔTC, ΔHDL, Δ trig, Δgly and ΔClCr as well as Wilcoxon non-parametric tests on the appearance of grade 2 diarrhoea.

Eur J Clin Pharmacol (2009) 65:667–678

Results

Patients

Forty-two patients were included in this treatment group of the COPHAR 2 ANRS–111 trial. However, one patient withdrew from the study, and one switched to another PI during the first week of treatment. We therefore obtained PK data from 40 patients (27 men, 13 women) with a median age of 36.5 years (range 20.0–59.0 years). Table 1 summarizes the main characteristics of the patient cohort.

Both of the approaches used to allocate the ethnic group provided corroborating results. Using the civic information we allocated 20 patients to the African group and 20 to the Caucasian group. Because information for all genotypes was missing for all genotypes, the Structure software allocated 19 patients to the Caucasian group and 20 to the African group. In the resulting two ethnic groups, Hardy–Weinberg proportions were respected for all polymorphisms under study, as shown in Table 2.

Indinavir pharmacokinetics

Two samples were missing, the trough and the 6 h concentrations, for two patients, and only the trough concentration was available for a second patient. Among the 155 samples, two indinavir plasma concentrations in one patient were below the LOQ (at 1 h and at trough), and these were discarded from further analysis. Figure 1a shows the plot of indinavir plasma concentrations at W2 versus time, revealing a high inter-individual variability.

The best error model was a proportional error model. The population estimates are displayed in Table 3. All of the relative standard errors (RSE) were below 25% with the exception of k_a and $\omega_{V/F}$ (around 30 and 60%, respectively). The inter-individual variance of k_a in this study was rather important (above 100%). The simulated median and the 90th interval are given in Fig. 2a together with all of the observed concentrations of indinavir. This graph provides good evidence of the adequacy of the model.

From that basic model, we first tested the effects of the covariates on the individual parameter estimates. Effects of age (P=0.03) and the ABCB1 exon 26 polymorphism (P=0.09) on Cl/F and of the Centers for Disease Control (CDC) classification (P=0.09) and the *CYP3A4**1B polymorphism (P=0.09) on k_a were found. Both ABCB1 exon 26 and the *CYP3A4**1B polymorphism variables were dichotomised in mutant homozygotes versus other genotypes. Following a forward selection based on LRT, the population model had *CYP3A4* effect on k_a (P=0.02) and an age effect on Cl/F (P =0.03). The age effect on clearance was withdrawn from the model after the backward selection based on the permutation test. In the final model, the absorption rate constant was decreased by 70% (P=0.04, LRT by permutation) in patients with the *1B*1B genotype for the *CYP3A4* allele:

$$k_a = 2.1 \times e^{-1.3 \times CYP3A4} \text{with} \begin{cases} \text{CYP3A4} = 0 \text{ for patients CYP3A4} * 1\text{A} * 1\text{A or CYP3A4} * 1\text{A} * 1\text{B} \\ \text{CYP3A4} = 1 \text{ for patients CYP3A4} * 1\text{B} * 1\text{B} \end{cases}.$$

Table 1 Characteristics of the patient cohort (n = 40)

Characteristics of the patients	Median (range)
Age (years)	36.5 (20.0–59.0)
BMI (kg/m^2)	22.6 (17.5–35.8)
Weight (kg)	68.0 (45.0–103.0)
Creatinine clearance (mmol/L)	95.4 (57.4–245.1)
Albumin (g/L)	38.4 (25.5–47.4)
Orosomucoid (g/L)	1.0 (0.5–2.9)
	Number of patients (%)
Dose (400/600/800 mg)	26 (65)/8 (20)/6 (15)
Coadministration of AZT/3TC (y/n)	33 (83)/ 7 (17)
Coinfection VHB/VHC (yes/no)[a]	7 (18)/32 (82)
Good adherence (yes/no)	15 (38)/25 (62)
Sex (male/female)	27 (68)/13 (32)
Ethnic group (African/Caucasian)	20 (50)/20 (50)
CDC classification for HIV infection (A or B/C)	30 (75)/10 (25)

BMI, Body mass index; AZT/3TC, zidovudine lamivudine combination; VHC/VHB, co-infection by hepatitis C or B; CDC Centers for Disease Control; HIV, human immunodeficiency virus

[a] Data on one patient are missing

Table 2 Distribution of the genetic polymorphisms within each ethnic group and Hardy–Weinberg P values

Genetic polymorphisms	Number of patients (%)	H-W P-value
African		
ABCB1 exon 26 (CC/CT/TT)	11 (55)/9 (45)/ 0 (0)	0.43
ABCB1 exon 21 (GG/GT/TT)	19 (95)/1 (5)/ 0 (0)	0.99
CYP3A5 (4*1/3*1/_2*1)	0 (0)/8 (40)/12 (60)	0.53
*CYP3A4**1B (*1A*1A/*1A*1B/*1B*1B)	9 (45)/8 (40)/3 (15)	0.86
Caucasian		
ABCB1 exon 26 (CC/CT/TT)	2 (12)/12 (70)/3 (18)	0.22
ABCB1 exon 21 (GG/GT/TT)	4 (21)/11 (58)/4 (21)	0.79
CYP3A5 (4*1/3*1/_2*1)	18 (100)/0 (0)/0 (0)	1
*CYP3A4**1B (*1A*1A/*1A*1B/*1B*1B)	0 (0)/3 (16)/16 (84)	0.93

H-W P value, Hardy–Weinberg P value according to the H–W proportions test

Data on all genotypes were missing for one patient; data on the ABCB1 exon 26 and *CYP3A4* genotypes were both missing for a second patient; data on the genotype for ABCB1 exon 26 were missing for a third patient

The population parameters of this final model and their RSE are given in Table 3 for the 38 patients with data available genotyping for *CYP3A4**1B polymorphism. The inter-individual variability for k_a decreased by 27% from the basic model with the incorporation of the covariate, and residual variability was 44.7%.

Ritonavir pharmacokinetics

For one patient, only data on the indinavir concentrations were available and there was no data on ritonavir concentration; consequently, we only analysed ritonavir data for 39 patients. The same five samples for indinavir mentioned in the preceding section were also missing. Among the 151 samples, two ritonavir plasma concentrations at 1 h and at trough in one patient and one concentration at 12 h in another patient were below the LOQ and were discarded. Observed plasma concentrations are given in Fig. 1b, and it should be noted that some patients showed high plasma concentrations (above 2000 ng/mL) for a dose of 100 mg b.i.d.

A proportional error model was selected. The population estimates are displayed in Table 4. All of the RSE were below 25% with the exception of k_a; this was partly attributable to the sparse design and to the $\omega_{Cl/F}$, as observed for the indinavir data. The VPC obtained with the basic model parameters estimates is given in Fig. 2b, together with the concentrations observed.

The results of the basic model evaluation were very satisfactory.

Effects of orosomucoid (P=0.03), albumin levels (P=0.04) and *CYP3A5* polymorphism (patients with two wild alleles at most vs. other genotypes, P=0.04) on Cl/F were

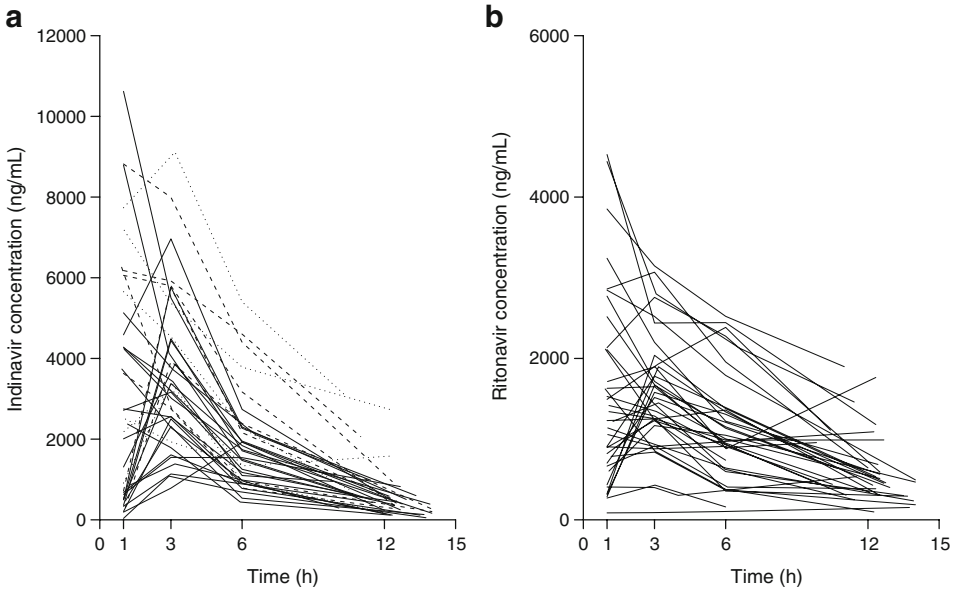


Fig. 1 Observed plasma indinavir concentration (**a**) and plasma ritonavir concentration (**b**) versus time in samples collected 2 weeks after treatment initiation, in 40 human immunodeficiency virus (HIV) naive-patients receiving indinavir plus 100 mg of ritonavir twice daily (b.i.d). In the indinavir plot, the *solid lines* correspond to an indinavir dose of 400 mg b.i.d., the *dashed lines* to 600 mg b.i.d. and the *dotted lines* to 800 mg b.i.d. Sampling times following drug administration were measured by the nurse. Concentrations were assumed at steady state, trough concentrations are those of samples taken following drug intake at sampling times deduced from the patient record

Table 3 Population pharmacokinetic parameters of indinavir for the basic and the final model: estimates and relative standard error

Parameters	Basic model (n=40)		Covariate model (n=38)	
	Estimates	RSE (%)	Estimates	RSE (%)
$k_a(h^{-1})$	1.3	33.7	2.1	44.1
$\beta_{k_a}^{CYP3A4}$	-	-	-1.3	42.0
Cl/F (L/h)	21.9	6.9	22.2	6.9
V/F (L)	93.9	8.2	97.3	9.3
ωk_a (%)	118.0	22.9	98.2	28.7
ωCl/F (%)	34.4	15.0	34.9	15.0
ωV/F (%)	19.3	66.8	21.6	57.8
σ(%)	44.5	8.9	44.7	8.6

k_a, absorption rate constant; Cl/F, oral clearance; V/F, oral volume of distribution; RSE, relative standard error

found on the individual parameters by the non-parametric tests, as were effects of HIV disease status (P=0.05) on $k_a$and creatinine clearance (P=0.1) on V/F. In the final model, an increase of 0.5 g/L in orosomucoid from the median (1 g/L) was associated with a clearance decrease of 28% (P=0.03, LRT by permutation):

$$Cl/F = 8.3 \times Orosomucoid^{-0.8}$$

The population parameters of this model and their RSE are given in Table 4.

Link between indinavir and ritonavir PK parameters

Four positive correlations between individual parameters of ritonavir and indinavir were found to be significant. There was a relationship between the indinavir and ritonavir absorption rate constant (r=0.4, P=0.005). Indinavir clearance was strongly correlated to ritonavir clearance (r=0.6, P <0.0001) and to a smaller degree to ritonavir volume of distribution (r=0.4, P <0.01), while indinavir volume of distribution was highly correlated to ritonavir volume of distribution (r=0.5, P < 0.002).

Concentrations link with short-term efficacy and safety

There was a significant decrease in viral load in the first 2 weeks of treatment, and a significant increase in total cholesterol, glycaemia and triglycerides in the first 4 weeks of treatment, as shown in Table 5.

The decrease in log viral load was significantly associated with higher indinavir AUC (r= −0.4, P=0.03) and C_{trough} (r= −0.4, P=0.03), as shown in Fig. 3. No significant difference in viral load decrease was found between the five patients with a C_{trough} below the lower limit of the therapeutic range and the 35 patients with a C_{trough} above this value.

Further, no significant relationship was found between indinavir nor ritonavir concentrations and safety measurements or grade 2 diarrhoea. No nephrolithiasis has been reported in the COPHAR 2–ANRS 111 trial, which has prevented us from analysing the link between concentrations and this adverse event associated with indinavir.

The genetic covariate kept in the final population PK model was the *CYP3A4*1B polymorphism categorised in two classes: *1B*1B versus other genotypes. Both C_{max} and increase in triglycerides were found to be significantly associated with the *CYP3A4*1B polymorphism, although the correlation was not significant. the C_{max} was significantly lower in patients homozygous for the *1B allele (median 3.6, range 2.2–5.2 ng/mL) than in the other groups (mean 4.4, range 2.2–8.3 ng/mL) (P=0.04), and the increase in triglycerides was also significantly smaller (mean 0.1, range −0.7 to 1.4 vs. mean 0.6, range −0.5 to 1.7 mmol/L, respectively; P=0.02), as illustrated by Fig. 4. In terms of the efficacy, no significant association was found between the *CYP3A4*1B*1B genotype and the C_{trough} or the log viral decrease.

The various doses of indinavir were not found to be associated with the *CYP3A4*1B polymorphism, short-term

Table 4 Population pharmacokinetic parameters of ritonavir for the basic and the final model: estimates and RSE

Parameters	Basic model (n=39)		Covariate model (n=39)	
	Estimates RSE (%)	Estimates RSE (%)	Estimates RSE (%)	Estimates RSE (%)
$k_a(h^{-1})$	2.4	98.7	2.2	93.5
Cl/F (L/h)	8.7	9.4	8.3	9.0
$\beta_{Cl}^{Orosomucoid}Z$	-	-	–0.8	46.5
V/F (L)	61.6	8.6	60.7	8.7
ωk_a (%)	357.7	21.9	346.8	21.5
ωCl/F (%)	55.9	12.2	52.4	12.4
ωV/F (%)	22.8	53.9	23.3	51.5
σ (%)	30.4	8.4	30.3	8.4

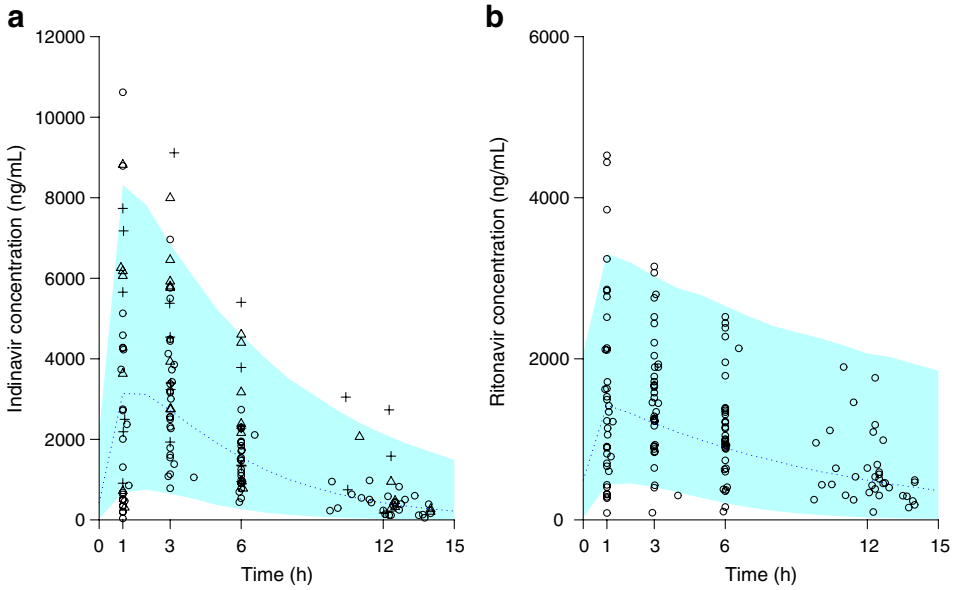


Fig. 2 Visual predictive check of the basic population pharmacokinetics (PK) model: comparison between the median (*line*) and the 90th interval (*shaded area*) predicted for 1000 simulated data sets and the observed concentrations of indinavir (**a**) and of ritonavir (**b**). Indinavir plot: *open circles* indinavir dose of 400 mg, *open triangles* indinavir dose of 600 mg, *crosses* indinavir dose of 800 mg

efficacy or safety, which negated its potential confounding effect.

Discussion

The PK of indinavir was analysed using a one-compartment model with first-order absorption and elimination at steady-state. The estimated clearance and volume of distribution were 22.2 L/h and 97.3 L, respectively, both of which are in the range of those obtained in previous studies [9, 11, 32]. In this study, ABCB1 exons 26 and 21 and the *CYP3A5**3 and *6 polymorphisms were not found to significantly influence the PK of indinavir: the absorption rate was 0.6 h^{-1} for *CYP3A4**1B*1B patients and 2.1 h^{-1} for *CYP3A4**1A*1A or *CYP3A4**1A*1B patients. The *CYP3A* enzymes are distributed in both hepatocytes and enterocytes [33] and their inhibition by ritonavir is well-documented [34–36]. In vivo, the genotype–phenotype correlation for *CYP3A4**1B remains a subject of debate [37–40]; however, *CYP3A4**1B has been related to increased transcription [41] in vitro. We hypothesised that in *CYP3A4**1B*1B patients, the ritonavir inhibition potency is lowered, leading to a higher first pass effect of indinavir, although this does not impact on its clearance. The potential confounding effect of the ethnic group was discarded, as this covariate was not significantly related to indinavir individual parameters in the sample. However, this finding is more relevant clinically in an African population given the extremely low frequency of the *CYP3A4**1B*1B genotype among Caucasians. The primary objective of the COPHAR2 study was not to assess the influence of genetic polymorphisms on indinavir PK, and the use of modelling has helped to circumvent the limited sample size of 40 patients in the study. In addition, most of

Table 5 Median and range of the studied short-term efficacy and safety measurements and of the change from baseline

Short-term efficacy and safety measurements	Baseline[a]	W2 or W4[b]	Difference from baseline	*P* value
Efficacy				
Log viral load (log copies/mL)	4.9 (3.4–6.3)	2.9 (1.8–4.1)	−1.8 (–2.8 to −0.5]	<0.001
Safety				
Total cholesterol (mmol/L)	4.3 (1.9–7.4)	5.0 (2.9–7.5)	0.8 (0.8–4.7)	<0.001
HDL cholesterol (mmol/L)	1.1 (0.5–1.8)	1.1 (0.4–2.1)	0.1 (– 0.7–1.0)	0.09
Glycaemia (mmol/L)	4.7 (3.4–6.0)	4.9 (2.8–7.1)	0.2 (−1.0 to 2.7)	0.013
Triglycerides (mmol/L)	1.0 (0.4–3.0)	1.4 (0.6–4.0)	0.4 (−0.7 to 1.7)	<0.001
Creatinine clearance (mL/min)	98.4 (62.0–195.7)	97.4 (62.8–252.0)	−1.0 (−38.0 to 56.4)	0.5

HDL, High-density lipoprotein

[a] Baseline = Day 0 for log viral load and week (W) 4 for safety

[b] Week 2 for log viral load and week 4 for safety

Eur J Clin Pharmacol (2009) 65:667–678

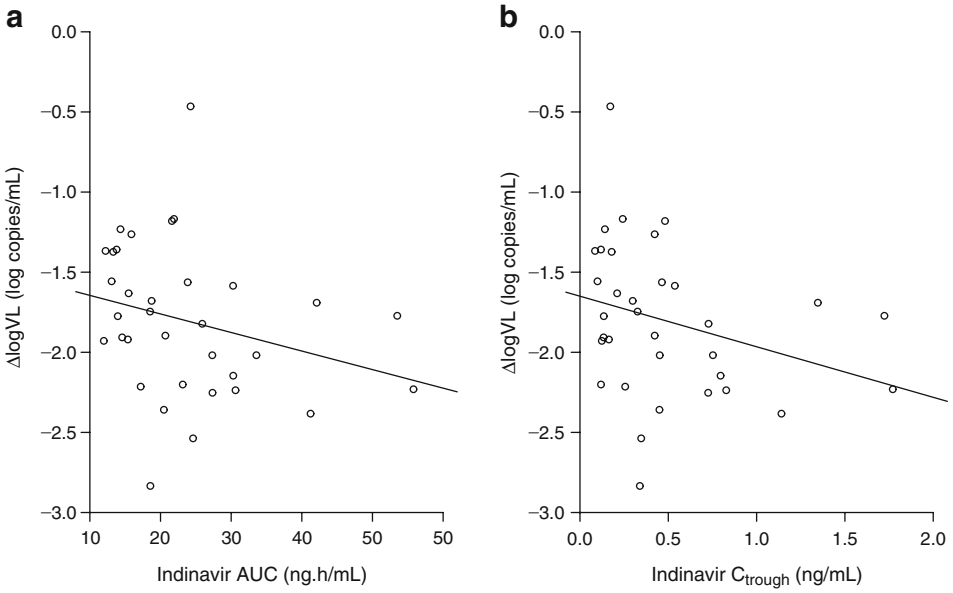


Fig. 3 Differences in log viral load (Δ logVL) observed between treatment initiation and week 2 versus area under the concentration–time curve (**a**) and trough plasma concentration of indinavir (**b**) predicted by the model

the tests in this study were performed as an exploratory step, and final inclusion in the model was based on permutation to cope with departure from the asymptotic assumption [30]. No evidence for a gender effect was found, as has been reported in a number of other studies on indinavir PK [9–11], but there were only 13 women in the present study. Dose has been found not to influence the PK of indinavir, and the use of ritonavir as a booster has been found to hide the dose non-linearity of indinavir [42]. We did not assess the impact of diet, as these data were not available, but patients were recommended to ingest the pills with food containing a sufficient amount of fats.

We also performed a population PK analysis of ritonavir concentrations. Ritonavir profiles were adequately described by a one-compartment model with first-order absorption and elimination processes, with estimates of the parameters being in good agreement with those of previous studies [6, 43, 44]. The estimated inter-individual variance for the absorption constant was singularly large. We found a negative relationship between ritonavir clearance and orosomucoid level in

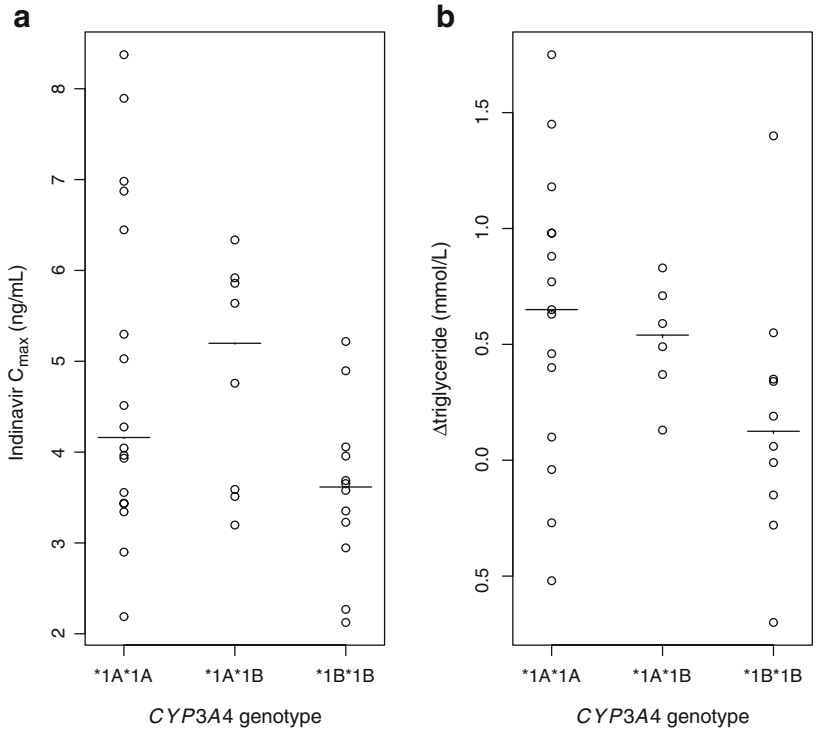


Fig. 4 Peak indinavir concentrations predicted by the model (**a**) and differences in triglycerides (*Δ triglyceride*) 4 weeks before and after treatment initiation (**b**) versus *CYP3A4* genotype. The *solid line* represents the median in each group

plasma. The affinity of ritonavir for orosomucoid protein as well as its impact on PI intracellular concentrations and efficacy has been described in both in vitro and in vivo studies [45–47]. In patients with high orosomucoid plasma levels, the decrease in the unbound fraction of ritonavir led to a lower clearance.

In the analysis of both PI, the few concentrations (1.3 and 2% for indinavir and ritonavir, respectively) below the LOQ were discarded. Using this approach, SAEM acquires a less important bias than it would with LOQ/2 [48]. There is no proper method in MONOLIX 2.1 to handle LOQ.

In the analysis of the link between indinavir and ritonavir concentrations, we chose not to include ritonavir as a covariate in the indinavir model, as performed in previous studies [10, 11]. Indeed, such parameterisation assumes a unidirectional influence of ritonavir on indinavir, which is not true. Ritonavir concentrations, when ritonavir is given with lopinavir, are lower than when ritonavir is given with indinavir [44]. We have instead emphasised the different levels of interaction between indinavir and ritonavir PK, especially at the absorption step, with the strong correlation between their absorption constant, but also in terms of bioavailability, as the oral clearances and volumes of distribution were highly correlated.

In order to properly model such an interaction between PI, a joint population analysis of concentrations of indinavir and ritonavir should be considered with correlated absorption constants and bioavailabilities.

We observed significant changes in viral load after 2 weeks of treatment, and we confirmed the association between high indinavir trough and mean concentrations and a greater decrease of viral load, which has already been described in PI-naive patients [49–51]. We did not find any relationship between *CYP3A*4*1B polymorphism and viral load decrease. We also observed a significant increase, after 4 weeks of treatment, of total cholesterol, glycaemia and triglycerides, as already reported [52], which was, however, not significantly related to indinavir concentrations at week 2. Ritonavir was found at singularly high levels in our study and is known to affect metabolic profiles, yet we found no evidence of an association between ritonavir levels and safety measurements. In patients homozygous for the *CYP3A*4*1B allele, the ritonavir-decreased inhibition on indinavir metabolism led to significantly lower indinavir C_{max} and appeared to impact at a metabolic level through a significantly lower increase in triglycerides in these patients.

Conclusion

We have developed and validated models for indinavir and ritonavir PK with reduced sampling in indinavir HAART patients. Both the average and trough concentrations were found to be predictors of the viral load decline. Only the *CYP3A*4*1B allele was found to influence indinavir absorption and biochemical safety, but no evidence was found of an impact of the five genetic polymorphisms studied on indinavir efficacy.

Acknowledgements Steering committee of the COPHAR2 ANRS-111 trial: principal investigators: D. Salmon- Céron, X. Duval, statistics: F. Mentré; other members: S. Auleley, M. Biour, M.J. Commoy, B. Diquet, C. Goujard, C. Katlama, C. Lascoux, M. Legrand, A. Métro, G. Peytavin, E. Rey, A.M. Taburet, J.M. Tre´luyer.

Safety committee: S. Auleley, M. Biour, A. Métro, C. Lascoux, D. Salmon-Céron. Pharmacological monitoring committee: X. Duval, E. Rey, J.M. Tréluyer. Independent committee: C. Rouzioux, C. Piketti, P. Flandre, M. Zenut, P. Marquet.

Clinical centers: Dr. Bentata, Dr. Mansouri, Mme Touam, Pr. Sereni, Dr. Lascoux, Dr. Pintado, Dr. Goujard, Mme Mole, Dr. Sellier, Dr. Bendenoun, Dr. Rami, Mme Parrinello, Dr. Jeantils, Mme Tassi, Pr. Vittecoq, Dr. Teicher, Mme Mallet, Pr. Dupont, Dr. Lahoulou, Soeur Azar, Pr. Rosembaum, Dr. Slama, Dr. Naï-Ighil Baakili, Dr. Courtial-Destembert, Pr. Vildé, Pr. Leport, Dr. Duval, Dr. Al Kaied, Pr. Salmon, Dr. Spiridon, Dr. Lesprit, Mme Chesnel, Pr. Katlama, Dr. Schneider, Mme Schoen, Pr. Molina, Dr. Ponscarme, Dr. Colin de verdière, Pr. Morlat, Dr. Bonarek, Dr. Joly, Dr. Ralaimazava, Mme Meridda, Mme Le Gac, P. Raffi, Dr. Allavena, Mme Hüe, Mme Sicot, Dr. Perré, Dr. Leautez, Dr. Aubry, Mme Suaud, Pr. Dellamonica, Dr. Rahelinirina, Pr. Michelet, Dr. Bouvier, Pr. Bazin, Dr. Goubin, Pr. May, Dr. Boyer, Pr. Rouveix, Dr. Dupont, Mme Berthé.

Pharmacological centers: Dr. Rey, Pr. Tréluyer, Dr. Abbara, Dr. Audoul, Dr. Tran, Dr. Sauvageon, Dr. Poirier, Dr. Taburet, Dr. Vincent, Dr. Aymard, Dr. Peytavin, Dr. Lamotte, Dr. Dailly, Dr. Garraffo, Dr. Lavrut, Dr. Mollimard, Dr. Titier, Dr. Tribut, Dr. Hulin, Dr. Huet, Dr. Delhotal, Dr. Hoizey.

Virological centers: Pr. Nicolas-Chanoine, Dr. Sousan, Pr. Dény, Dr. Baazia, Dr. Alloui, Pr. Brun-Vézinet, Dr. Chams, Pr. Fleury, Dr. Pellegrin, Dr. Garrigue, Pr. Fremut, Dr. Vabret, Pr. Lebon, Dr. Krivine, Pr. Calvez, Dr. Gourlain, Dr. Amellal, Pr. Bouvier-Alias, Pr. Norman, Dr. Idri, Pr. Chambreuil, Dr. Poirier, Pr. Mazeron, Pr. Le Faou, Dr. Vénard, Pr. Billaudel, Dr. Ferre, Pr. Rouzioux, Dr. Burgard, Pr. Lefevre, Dr. Cottalorda, Pr. Dussaix, Dr. Bensidhoum, Pr. Colimon, Dr. Ruffault, Dr. Maillard, Pr. Morinet, Dr. Palmer, Pr. Nicolas, Dr. Zalta.

Monitoring: S. Auleley, E. Marcault, F. Mentré. Statistics: E. Bougen, F. Mentré, X. Panhard.

The authors thank the study participants and the Agence de Recherche Nationale sur le SIDA (ANRS, Essai 111) for financial support.

We also acknowledge Dr. Emmanuelle Comets, Pr. Marc Lavielle and Dr. Emmanuelle Génin for their valuable advices during the analysis.

J. Bertrand was supported by a grant from Servier Research Group, France.

References

1. Solas C, Lafeuillade A, Halfon P et al (2003) Discrepancies between protease inhibitor concentrations and viral load in reservoirs and sanctuary sites in human immunodeficiency virus-infected patients. Antimicrob Agents Chemother 47:238–243
2. Letendre S, Marquie-Beck J, Capparelli E et al (2008) Validation of the CNS penetration-effectiveness rank for quantifying antiretroviral penetration into the central nervous system. Arch Neurol 65:65–70

3. Collin F, Chêne G, Retout S, Peytavin G, Salmon D, Bouvet E, Raffi F, Garraffo R, Mentré F, Duval X, ANRS CO8 Aproco-Copilote Study Group (2007) Indinavir trough concentration as a determinant of early nephrolithiasis in HIV-1-infected adults. Ther Drug Monit 29:164–170
4. Flexner C (1998) HIV-protease inhibitors. N Engl J Med 338:1281–1292
5. Von Moltke LL, Greenblatt DJ, Grassi JM et al (1998) Protease inhibitors as inhibitors of human cytochromes P450: high risk associated with ritonavir. J Clin Pharmacol 38:106–111
6. Hsu A, Granneman GR, Bertz RJ (1998) Ritonavir. Clinical pharmacokinetics and interactions with other anti-HIV agents. Clin Pharmacokinet 35:275–291
7. Soriano V, Garcya-Gasco P, Vispo E, Ruiz-Sancho A, Blanco F, Martyn-Carbonero L, Rodryguez-Novoa S, Morello J, de Mendoza C, Rivas P, Barreiro P, Gonzalez-Lahoz J (2008) Efficacy and safety of replacing lopinavir with atazanavir in HIV-infected patients with undetectable plasma viraemia: final results of the SLOAT trial. J Antimicrob Chemother 61:200–205
8. Collot-Teixeira S, De Lorenzo F, Waters L, Fletcher C, Back D, Mandalia S, Pozniak A, Yilmaz S, McGregor JL, Gazzard B, Boffito M (2009) Impact of different low-dose ritonavir regimens on lipids, CD36, and adipophilin expression. Clin Pharmacol Ther 85(4):375–378
9. Csajka C, Marzolini C, Fattinger K et al (2004) Population pharmacokinetics of indinavir in patients infected with human immunodeficiency virus. Antimicrob Agents Chemother 48:3226–32
10. Brendel K, Legrand M, Taburet A et al (2005) Population pharmacokinetic analysis of indinavir in hiv-infected patient treated with a stable antiretroviral therapy. Fundam Clin Pharmacol 19:373–383
11. Kappelhoff BS, Huitema ADR, Sankatsing SUC et al (2005) Population pharmacokinetics of indinavir alone and in combination with ritonavir in HIV-1-infected patients. Br J Clin Pharmacol 60:276–286
12. Solas C, Simon N, Drogoul MP et al (2007) Minimal effect of MDR1 and CYP3A5 genetic polymorphisms on the pharmacokinetics of indinavir in HIV-infected patients. Br J Clin Pharmacol 64:353–362
13. Verstuyft C, Marcellin F, Morand-Joubert L et al (2005) Absence of association between MDR1 genetic polymorphisms, indinavir pharmacokinetics and response to highly active antiretroviral therapy. AIDS 19:2127–2131
14. Anderson PL, Lamba J, Aquilante CL et al (2006) Pharmacogenetic characteristics of indinavir, zidovudine, and lamivudine therapy in HIV-infected adults: a pilot study. J Acquir Immune Defic Syndr 42:441–449
15. Lamba JK, Lin YS, Schuetz EG et al (2002) Genetic contribution to variable human CYP3A-mediated metabolism. Adv Drug Deliver Rev 54:1271–1294
16. Lichterfeld M, Nischalke HD, Bergmann F et al (2002) Long-term efficacy and safety of ritonavir/indinavir at 400/400 mg twice a day in combination with two nucleoside reverse transcriptase inhibitors as first line antiretroviral therapy. HIV Med 3:37–43
17. Duval X, Mentre F, Lamotte C et al (2005) Indinavir plasma concentration and adherence score are codeterminant of early virologic response in HIV-infected patients of the APROCO cohort. Ther Drug Monit 27:63–70
18. Dieleman JP, Gyssens IC, van der Ende ME et al (1999) Urological complaints in relation to indinavir plasma concentrations in HIV-infected patients. AIDS 13:473–478
19. Duval X, Mentré F, Rey E et al (2009) Benefit of therapeutic drug monitoring of protease inhibitors in HIV-infected patients depends on PI used in HAART regimen—ANRS 111 trial. Fundam Clin Pharmacol (in press)
20. Hirt D, Mentré F, Tran A et al (2008) Effect of CYP2C19 polymorphism on nelfinavir to M8 biotransformation in HIV patients. Br J Clin Pharmacol 65:548–557
21. Carrieri P, Cailleton V, Le Moing V et al (2001) The dynamic of adherence to highly active antiretroviral therapy: results from the french national APROCO Cohort. J Acq Immun Def Synd 28:232–239
22. Cascorbi I, Gerloff T, Johne A et al (2001) Frequency of single nucleotide polymorphisms in the P-glycoprotein drug transporter MDR1 gene in white subjects. Clin Pharmacol Ther 69:169–174
23. Dally H, Edler L, Jager B et al (2003) The CYP3A4*1B allele increases risk for small cell lung cancer: effect of gender and smoking dose. Pharmacogenetics 13:607–618
24. Crow J (1999) Hardy, Weinberg and language impediments. Genetics 152:821–825
25. Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
26. Kuhn E, Lavielle M (2005) Maximum likelihood estimation in nonlinear mixed effects models. Comput Stat Data Anal 49:1020–1038
27. Lavielle M (2008) MONOLIX (MOdèles NOn LInéaires à effets miXtes). MONOLIX group, Orsay, France. Available at: http://software.monolix.org/index.php
28. Lavielle M, Mentré F (2007) Estimation of population pharmacokinetic parameters of saquinavir in HIV patients with the MONOLIX software. J Pharmacokinet Pharmacodyn 34:229–249
29. Gelman A, Carlin JB, Stern HS et al (1995) Bayesian data analysis. Chapman & Hall, London
30. Bertrand J, Comets E, Mentré F (2008) Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. J Biopharm Stat 18:1084–1102
31. Manly B (1998) Randomization. bootstrap and Monte Carlo methods in biology. Chapman & Hall, London
32. Goujard C, Legrand M, Panhard X et al (2005) High variability of indinavir and nelfinavir pharmacokinetics in HIV-infected patients with a sustained virological response on highly active antiretroviral therapy. Clin Pharmacokinet 44:1267–1278
33. Fuhr U, Jetter A, Kirchheiner J (2007) Appropriate phenotyping procedures for drug metabolizing enzymes and transporters in humans and their simultaneous use in the "cocktail" approach. Clin Pharmacol Ther 81:270–283
34. Ikezoe T, Hisatake Y, Takeuchi T et al (2004) HIV-1 protease inhibitor, ritonavir: a potent inhibitor of CYP3A4, enhanced the anticancer effects of docetaxel in androgen-independent prostate cancer cells in vitro and in vivo. Cancer Res 64:7426–7431
35. Zhou S, Yung Chan S, Cher Goh B et al (2005) Mechanism-based inhibition of cytochrome P450 3A4 by therapeutic drugs. Clin Pharmacokinet 44:279–304
36. Rittweger M, Arastéh K (2007) Clinical pharmacokinetics of darunavir. Clin Pharmacokinet 46:739–756
37. Spurdle A, Goodwin B, Hodgson E et al (2002) The CYP3A4*1B polymorphism has no functional significance and is not associated with risk of breast or ovarian cancer. Pharmacogenetics 12:355–366
38. He P, Court MH, Greenblatt DJ et al (2005) Genotype-phenotype associations of cytochrome P450 3A4 and 3A5 polymorphism with midazolam clearance in vivo. Clin Pharmacol Ther 77:373–387
39. Hesselink DA, van Schaik RH, van der Heiden IP et al (2003) Genetic polymorphisms of the CYP3A4, CYP3A5, and MDR-1 genes and pharmacokinetics of the calcineurin inhibitors cyclosporine and tacrolimus. Clin Pharmacol Ther 74:245–254

40. Tran A, Jullien V, Alexandre J et al (2006) Pharmacokinetics and toxicity of docetaxel: role of CYP3A, MDR1, and GST polymorphisms. Clin Pharmacol Ther 79:570–580
41. Amirimani B, Ning B, Deitz AC et al (2003) Increased transcriptional activity of the CYP3A4*1B promoter variant. Environ Mol Mutagen 42:299–305
42. Yeh K, Stone J, Carides A et al (1999) Simultaneous investigation of indinavir nonlinear pharmacokinetics and bioavailability in healthy volunteers using stable isotope labeling technique: study design and model-independent data analysis. J Pharm Sci 88:568–573
43. Lu JF, Blaschke TF, Flexner C et al. (2002) Model-based analysis of the pharmacokinetic interactions between ritonavir, nelfinavir and saquinavir after simultaneous and staggered oral administration. Drug Metab Dispos 30:1455–1461
44. Kappelhoff BS, Huitema ADR, Crommentuyn KML et al (2005) Development and validation of a population pharmacokinetic model for ritonavir used as a booster or as an antiviral agent in HIV-1-infected patients. Br J Clin Pharmacol 59:174–182
45. Schön A, del Mar Ingaramo M, Freire E (2003) The binding of HIV-1 protease inhibitors to human serum proteins. Biophys Chem 105:221–230
46. Jones K, Hoggard PG, Khoo S et al (2001) Effect of alpha1-acid glycoprotein on the intracellular accumulation of the HIV protease inhibitors saquinavir, ritonavir and indinavir in vitro. Br J Clin Pharmacol 51:99–102
47. Zhang X, Schooley R, Gerber J (1999) The effect of increasing alpha1-acid glycoprotein concentration on the antiviral efficacy of human immunodeficiency virus protease inhibitors. J Infect Dis 180:1833–1837
48. Samson A, Lavielle M, Mentré F (2006) Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: Application to HIV dynamics model. Comput Stat Data Anal 51:1562–1574
49. Fletcher CV (1999) Pharmacologic considerations for therapeutic success with antiretroviral agents. Ann Pharmacother 33:989–995
50. Burger DM, Hugen PWH, Aarnoutse RE et al (2001) A retrospective, cohort-based survey of patients using twice-daily indinavir+ritonavir combinations: pharmacokinetics, safety and efficacy. J Acquir Immune Defic Syndr 26:218–224
51. Rayner CR, Galbraith KJ, Marriott JL et al (2002) A critical evaluation of the therapeutic range of indinavir. Ann Pharmacother 36:1230–1237
52. Behrens G, Dejam A, Schmidt H et al (1999) Impaired glucose tolerance, beta cell function, and lipid metabolism: HIV patients under treatment with protease inhibitors. AIDS 13:F63–F70

3.2 Détection d'un effet gène dans un modèle complexe parent-métabolite

3.2.1 Résumé

Nous avons analysé les concentrations d'un antipsychotique recueillies dans un essai de phase II comparatif réalisé au cours de son développement par les laboratoires pharmaceutiques Servier.

Les concentrations de la molécule parent et de son métabolite actif ont été mesurées à quatre temps de prélèvements chez quatre-vingt-dix-sept patients après quatre semaines de traitement et chez soixante-onze patients après huit semaines. Nous avons réalisé une modélisation conjointe des concentrations de produit parent et de métabolite aux deux occasions. Dans le choix du modèle structural, nous avons intégré les connaissances sur le produit acquises lors des précédentes études de phase I. Nous avons comparé différents modèles plus ou moins complexes (mécanisme d'interconversion, effet de premier passage) et nous avons étudié la paramétrisation de ces modèles afin de gérer les questions d'identifiabilité.

Les données ont été analysées avec les algorithmes d'estimation FOCE et SAEM. Des solutions analytiques (SA) des modèles structuraux ont été dérivées afin de comparer les résultats suivant que le modèle était mis en œuvre avec un système d'équations différentielles ordinaires (ODE) ou avec une SA. En effet, les modèles complexes sont souvent analysés en ODE parce que déterminer une SA n'est pas toujours aisé. Nous avons voulu utiliser dans ce cadre l'extension pour les systèmes d'ODE, MLXTRAN, récemment implémentée dans le logiciel MONOLIX et comparer ses performances à celles du logiciel référent en pharmacocinétique de population, NONMEM. La sélection du modèle structural était basée sur le BIC. Pour chaque paramètre du modèle, les variabilités inter-sujets (IIV) et inter-occasions (IOV) ont été estimées.

L'effet des polymorphismes du *CYP* 2D6 et du *CYP* 2C19 a été exploré sur les différents paramètres du modèle structural sélectionné. Comme les concentrations ont été recueillies en plusieurs occasions, nous avons pu calculer pour tous les paramètres le coefficient de la part génétique de la variabilité $R_{GC} = 1 - \frac{IOV}{IIV}$ tel que défini par Ozdemir et al. (2000). Ce coefficient est d'autant plus proche de 1 que la variabilité du paramètre est susceptible d'être expliquée par des variations génétiques. La stratégie de sélection adoptée dans cette étude a consisté en une première exploration sur les EBE des paramètres entre les génotypes avec un modèle linéaire à effet mixtes. Puis les covariables significatives à cette étape étaient incluses dans le modèle de population par sélection ascendante basée sur le test de Wald. Comme les génotypes des 2 polymorphismes à l'étude était très déséquilibrés, les p-values des covariables restant dans le modèle final ont été calculées par permutation. Le modèle final a été validé par des graphiques appelés *visual predictive check* où les observations sont superposées à la médiane et l'intervalle de confiance à 90%

prédits par le modèle (Karlsson et Holford, 2008).

Pour le modèle le plus complexe, le temps de calcul était divisé par 30 lors du passage d'ODE en SA. Les estimations des paramètres étaient proches suivant que le modèle était mis en œuvre en ODE ou SA, à l'exception des constantes d'absorption avec MONOLIX. Le protocole, peu informatif dans la phase d'absorption, pourrait expliquer ce comportement. De plus, MONOLIX pour l'analyse de données à l'équilibre avec un modèle en ODE requiert l'ajout de doses fictives simulant l'atteinte de cet état. Un nombre insuffisant de doses pourrait alors expliquer ces disparités dans les constantes d'absorption.

Le même modèle structural a été sélectionné avec les algorithmes FOCE et SAEM en ODE et SA. Ce modèle inclut un mécanisme d'interconversion, un effet de premier passage avec fractionnement de la dose et la même vitesse d'absorption pour les deux molécules. Le paramètre avec la plus grande valeur de R_{GC} était la constante d'élimination du métabolite, et dans le modèle final, un effet du polymorphisme du gène *CYP* 2D6 a été mis en évidence sur ce paramètre.

Dans ce travail, nous avons pu modéliser la pharmacocinétique conjointe du produit parent et de son métabolite actif, et montrer l'effet du polymorphisme du gène *CYP* 2D6 sur la clairance d'élimination du métabolite ainsi que la répercussion sur la pharmacocinétique du produit parent, à travers le mécanisme d'interconversion.

L'ensemble de ce travail est décrit dans un article en préparation pour la revue *The American Association of Pharmaceutical Scientists Journal*.

3.2.2 Article 5 (en préparation)

Development of a complex parent-metabolite joint model and testing the influence of genetic covariates

Julie Bertrand[1], Céline M. Laffont[2], Emmanuelle Comets[1], Marylore Chenel[3], France Mentré[1]

[1] UMR738, INSERM, Paris, France; Université Paris Diderot, Paris, France
[2] UMR181, Physiologie et Toxicologie Expérimentales INRA, ENVT, Toulouse, France
[3] Institut de Recherches Internationales Servier, Courbevoie, France

Running head: Test for genetic effect on parent and metabolite PK

Corresponding author: Julie Bertrand, UMR 738, INSERM, Université Paris Diderot, 16 rue Henri Huchard, 75018 Paris, France; telephone: 33 (0)1 57 27 75 39; fax: 33 (0)1 57 27 75 21; e-mail: julie.bertrand@inserm.fr

ABSTRACT

We investigated the influence of single nucleotide polymorphisms on the pharmacokinetics of an antipsychotic agent under development and its active metabolite. The plasma concentrations of 101 patients were analysed on two occasions after 4 and 8 weeks of treatment at 1, 3, 6 and 24 hours (trough) following once a day administration. Concentrations were measured for both the parent drug and its active metabolite. For each patient, genotypes were obtained for *CYP* 2D6 and *CYP* 2C19 polymorphisms. We modelled jointly the concentrations of the antipsychotic and its active metabolite using FOCE with interaction in NONMEM version V and VI and SAEM in MONOLIX version 2.4. Four different structural models were compared encoded in a system of ordinary differential equations (ODE) and using a closed form solution (CF). The same structural model was selected using both algorithms, using either ODE or CF. The final model included two compartments with an back-transformation mechanism and a first-pass effect with a dose split. The volumes of the parent drug and the metabolite were set to be equal and the decreased bioavailability with the dose was taken into account. Between and within subject variabilities were estimated on all parameters with the exception of two. We performed a covariate screening on the EBE followed by an ascending selection based on Wald test. The clearance of elimination through other processes than the back-transformation of the metabolite was decreased by 47% in patients *CYP* 2D6 poor metabolizers (p-value=0.009, Wald test by permutation).

Key words: Nonlinear mixed effects modeling, Back-transformation mechanism, First-pass metabolism, Genetic covariate, Estimation algorithms

INTRODUCTION

During its development, an innovative antipsychotic agent from SERVIER research was shown to have a mainly metabolic elimination. Less than 1% of the parent drug and approximately 10% of the dose for the metabolite were recovered unchanged in urine. The parent drug was mainly metabolised through hydrolytic cleavage by hydrolases (amidases, deacetylases), as assessed in human microsomes which led to its active metabolite. To a lesser extent, the parent drug also underwent oxidative reactions involving CYP 3A4 and CYP 2C19. The active metabolite was itself metabolised by cytochrome P450s (mainly CYP 2D6) and also by N-acetyltransferases into the parent drug in an back-transformation process.

Both CYP 2C19 and CYP 2D6 are encoded by highly polymorphic genes and these polymorphisms are known to have an impact on the course of many therapeutic drugs [1]. Therefore in the phase II study, concentrations profiles were collected in patients at two occasions for the parent drug and its metabolite along with genotypes for 5 polymorphisms of the *CYP* 2D6 gene and 2 polymorphisms of the *CYP* 2C19 gene.

The aim of this work was to develop a joint population pharmacokinetic (PK) model for the parent drug and its active metabolite after oral administration in patients based on the data of phase II study and to analyse the genetic effect of *CYP* 2D6 and *CYP* 2C19 polymorphisms leading to extensive metabolizers (EM) and poor metabolizers (PM).

Nowadays, several approaches are proposed to investigate the effects of genetic polymorphisms in the literature. In previous works, we have shown through simulations that asymptotic tests require a correction for type I error inflation on designs with unbalanced genotypes and/or including small number of subjects [2, 3]. However, this slight inflation can be handled by randomisation tests [4]. Such computing intensive approaches require a powerful and stable estimation method, all the more when the structural and variability models become complex.

A secondary achievement of this work was the building of this model using the FOCE with interaction (FOCE-I) algorithm implemented in NONMEM [5] and the SAEM algorithm implemented in MONOLIX [6] in parallel, with models encoded both in ordinary differential equations (ODE) system and closed form (CF) solutions.

MATERIALS AND METHODS

Pharmacokinetic study

Data came from a pilot, phase II, international, multicentre, randomized, double-blind, parallel-group, controlled study with therapeutic benefit. Caucasian patients were randomly allocated to 4 groups of treatment receiving for 8 weeks either 5, 10 and 20 mg of the novel antipsychotic or the gold standard medication.

For each patient, four samples were taken: one prior to drug administration plus three at 1, 3 and 6 h after drug administration. The four samples were collected on two occasions: four and eight weeks after treatment initiation (W4 and W8). Plasma concentrations were assumed to be at steady state with trough concentrations considered to be observed at 14 h following the drug intake, using the delay reported by the patient from previous dosing. On the day of the sampling, the date and exact time of the drug administration and of each blood sampling were reported along with the date and exact time of the previous drug administration.

Here, we present the analysis of concentration-time profiles from patients treated with the novel antipsychotic only (excluding patients treated with the reference medication).

Concentration measurements and genetic polymorphisms

The plasma concentrations of the antipsychotic and its active metabolite were determined using a validated method involving solid phase extraction followed by reverse phase Liquid Chromatography with Mass Spectrometry - Mass Spectrometry detection (LC/MS-MS).

Blood samples were taken at the selection visit in order to determine the patient's genotype of *CYP* 2D6*3, *4, *6, *7, and *8 alleles and of *CYP* 2C19*2 and *3 alleles. All samples were stored and blood samples of included patients were sent on a regular basis to a central laboratory for analysis.

Joint modelling of the pharmacokinetics of the parent drug and its active metabolite

Below limit of quantification data and non reliable concentration data given the time or the dosing information were treated as missing data and were excluded from the analysis.

In a first step, the structural model was determined on data collected at W4 only. Four different structural models were investigated depending on the addition of i) a back-transformation mechanism, ii) a first-pass effect and iii) a dose fractionation (i. e. the first-pass effect includes a dose split: a fraction F_p of the dose leads to parent drug and a fraction 1-F_p leads to its active metabolite, prior to reach the plasma). Figure 1 regroups the structural representations of the four models along with the definition of all the parameters. Absorption and elimination were modelled as linear processes. For identifiability purposes, the volumes of both the parent drug and its active metabolite were set equal, since both molecules have similar molecular mass and physicochemical properties. Concentrations were expressed in nmol/L for the joint modelling of the parent drug and its metabolite which molecular weights were respectively, 319.4 g/mol and 361.4 g/mol.

The models were encoded in a system of ODE and using a corresponding CF solution that was derived using the Laplace transform approach [7, 8] (see appendix). For the models encoded with ODE system, data were fitted using the NONMEM software version V [5] and MLXTRAN

in MONOLIX version 2.4 [6]. For the models encoded in CF solution, data were fitted using the PRED routine in the NONMEM software version VI and the model building function of MONOLIX version 2.4.

The selection of the structural model was performed using the Bayesian information criteria [9]: $BIC = -2L + P_{pop}log(N)$, where L is the loglikelihood of the model, P_{pop} is the number of model parameters which includes the fixed effects and the variance components and N the number of subjects. For NONMEM, we retrieved $-2L$ for each model by adding $n_{tot} \times log2\pi$ to the objective function estimate, with n_{tot} the total number of observations. BIC allows to compare both nested and non nested models such as those with and without dose fractionation.

On data at W4, an exponential model was used to model between subject variability (BSV), with random effects assumed to follow a normal distribution. A full between subject variance matrix was tested *versus* a diagonal variance matrix where an additional parameter f was added which mean was fixed to 1 and which between subject variance was estimated. This additional parameter was bound to capture clearance and volume correlations. Variance parameters nullity was tested using likelihood ratio test (LRT) with a mixture of χ^2 as reference. Additive, proportional and combined error models were tested using LRT and goodness-of-fit plots (population and individual weighted residuals *vs* predicted concentrations and *vs* time).

In a second step, data at W8 were added to the data set and within subject variances were added on parameters with non null between subject variance.

The formula to obtain the area under the concentration-time curve (AUC) and half-life for both the parent drug and the metabolite were derived from the CF solution and based on [10, 11]. The proportion of dose that underwent back-transformation was computed as well.

Assessment of genetic covariates

The effects of the *CYP* 2D6 and *CYP* 2C19 polymorphisms were evaluated from the basic model. Both genetic covariates were analysed by means of a phenotypic binary categorisation, PM *versus* EM. The classification was performed as follows: two rare allele carriers were classified as PM [12].

The genetic component of variability (R_{GC}) for each model parameter was computed as 1 minus the ratio of the within subject variability (WSV) over the BSV. This component gets closer to 1 when the BSV for the parameter under study is larger than the WSV, so that the variability for this parameter is more likely to be explained by genetic covariate [13].

Empirical Bayes estimates (EBE) of each individual PK parameter were derived from the basic model at both occasions. The effects of genetic covariates on these EBE were tested in univariate analysis using linear mixed effect regression (lme function in R [14]). Genetic covariates which were found to have a significant effect (p-value<0.1) on EBE, were selected for inclusion in the population model. A forward selection of these covariates was then performed using a Wald test with a significance threshold at 5%. In order to correct the inflation of the Wald type I error shown

to occur on unevenly distributed genotypes [3], final p-values were assessed using permutations [4].

More specifically, 1000 data sets were generated by permuting the rows of the covariate matrix from the original data set. For each covariate, one Wald statistic, W^{obs}, was estimated from the original data and one Wald statistic, W^{perm} was estimated from each of the 1000 data sets. Thus, we obtained j=1,...,1000 W^{perm_j} . The permutation p-value was the proportion : (card($W^{perm_j} \geq W^{obs}$)+1)/(1000+1).

Validation was realized using visual predictive check plots, with the 90% confidence interval and median retrieved from 250 data sets simulated using the covariate model overlaid to the observed concentrations [15].

As in psychopharmacology, therapeutic drug monitoring studies have generally concentrated on keeping long-term exposure to the minimal effective blood concentration [16]. We simulated in each genotype the range of steady-state trough concentration predicted for 1000 individuals using the covariate model for both molecules with a 10 mg dose, given the total variability of the random effects without measurement error. This was performed in order to illustrate the extent of the effect associated to the genetic covariates.

RESULTS

Data

After data cleaning, 713 concentrations were measured at W4 in 97 patients among whom 81 had a complete profile of four samples. At W8, 539 concentrations were measured in 71 patients among whom 51 had a complete profile. Sixty seven patients had concentration-time profiles in both W4 and W8. The data set combining the observed profiles at both occasions contained 101 patients with a total of 1252 observations. Thirty five patients had a dose of 5 mg, 31 a dose of 10 mg and 35 a dose of 20 mg. Figure 2 displays the observed concentrations for the parent drug (a) and the metabolite (b) *versus* the time at W4 (left) and W8 (right), revealing the larger accumulation of the metabolite with regard to the parent drug.

Two patients had no genotype information and respectively 12 and 2 patients were classified as *CYP* 2D6 and *CYP* 2C19 PM. Three patients *CYP* 2D6 PM received a dose of 5 mg, 5 received a dose of 10 mg and 2 received a dose of 20 mg. Within the 20 mg group, one patient *CYP* 2D6 PM had no concentrations at W8. One patient *CYP* 2C19 PM received a dose of 5 mg and had concentrations in both occasions, the other received a dose of 20 mg and had concentrations only at W8.

Structural model

Convergence was achieved for all models with both algorithms and coding conditions. Considering the more complex model, computing times dramatically decreased from 2h30 to less than 3 min using ODE and CF respectively for both softwares. With NONMEM, the estimation variance step could not be achieved with any of the models on ODE or CF, with the exception of the model with first-pass effect but no dose fractionation encoded in CF. Whereas with MONOLIX, standard errors were obtained for all models using both codings. However, setting some between or within subject variances to 0 would have probably allow NONMEM to provide standard errors.

Table I reports the BIC along with the fixed effect estimates for the four structural models and a fifth model where absorption rates for the parent drug and the metabolite were set equal, using both software and codings. The ODE and CF solution estimates obtained with NONMEM were always quite close, while larger discrepancies were observed between ODE and CF estimates obtained with MONOLIX. Across models, BIC estimates tended to be lower when fitting with the SAEM algorithm and CF solution. The model selection was similar using either NONMEM or MONOLIX and either ODE or CF.

The BIC was decreased by 150 units from the model without to the model with the back-transformation mechanism. The model with back-transformation mechanism and first-pass effect had a BIC lower by 130 units than the model with back-transformation mechanism but no first-pass effect. The model with back-transformation and first-pass effect including dose fractionation had a BIC lower by 40 units than the model with back-transformation and first-pass effect without dose fractionation. Finally, the model with back-transformation and first-pass effect including dose fractionation with similar absorption rates for the antipsychotic and its metabolite had a BIC lower by 20 units than the same model with two different absorption rate estimates. The latter situation was investigated as the estimates of both absorption rates were close and the standard error coefficient variation for the metabolite parameter was quite high (about 90%). Thus, the selected model was the one including the back-transformation mechanism, the first-pass effect with dose fractionation and similar absorption rates for the parent drug and its active metabolite.

Given the shorter computing time and in order to obtain standard error estimates, the MONOLIX software with the model encoded in close form solution was used in the following.

Variability model

A diagonal BSV matrix was selected with variances on all parameters including the additional parameter f, but the fraction of dose that escapes first-pass effect F_p and the metabolism clearance CL_{mp} A linear dose effect was added on f with 10 mg as the reference dose, leading to a 15% increase and 25% decrease in f for 5 mg and 20 mg dose respectively. A proportional error model was selected for both the parent drug and its metabolite, with estimates for the residual variabilities below 30% and 10%, respectively. Within subject variances were estimated for all parameters with

non null between subject variances. The population parameters estimates of this basic model and their relative standard error (RSE) are given in Table II.

Secondary pharmacokinetic parameters

Using the population parameter estimates, the AUC for the parent drug were about 1.7, 3.0 and 4.5 mg.h/L for a dose of 5, 10 or 20 mg respectively, when for the metabolite AUC were about 9.6, 16.6 and 24.9 mg.h/L. The half-lives of the two molecules were 8.8 and 32.4 h. An estimated 8% of the dose underwent back-transformation.

Covariate model

R_{GC} estimates (i.e. 1-WSV/BSV for each parameter) were between 40-50% for all parameters, with exception of K_{ap} (and thus K_{am}) for which it was close to 0 and CL_{mo} for which it was equal to 74%. Thus, CL_{mo} appeared to be the parameter the most likely to be related to genetics.

In univariate analysis, effects of *CYP* 2D6 polymorphisms on CL_{po} (p-value=0.03) and on CL_{mo} (p-value=0.005) were found. After inclusion in the model, only the *CYP* 2D6 effect on CL_{mo} remained (p-value= 0,005). In this final model, CL_{mo} was decreased by 47% (p-value=0.009, Wald test by permutation) in *CYP* 2D6 PM patients. The population parameters of the covariate model and their RSE are given in Table II, for the 99 patients with available genotyping for *CYP* 2D6 polymorphisms. The BSV for CL_{mo} decreased of 3% with the incorporation of the covariate, but there were only 12 *CYP* 2D6 PM patients.

Figure 3 displays the VPC plots with the 90% interval and the median predicted from the covariate model overlaid to the observed concentrations for the parent drug (top) and the metabolite (bottom) for a dose of 5 mg (a), 10 mg (b) and 20 mg (c). The predictions from the model describe adequately well the observed concentrations of both molecules for the three doses, with exception of an over prediction of the variability on the metabolite concentrations observed for the 10 mg dose at W8.

Figure 4 displays the through concentrations distribution predicted by the covariate model, for the parent drug (a) and the metabolite (b) in the two groups of *CYP* 2D6 EM and PM patients for a dose of 10 mg. Median trough concentration increased by 1.5-fold in *CYP* 2D6 PM patients compared to *CYP* 2D6 PM patients for the parent drug and by 1.8-fold for the metabolite. However, for both molecules the remaining BSV was quite large.

DISCUSSION

In the present work, we compared five different structural models to describe the concentration-time profiles of a novel antipsychotic and its active metabolite collected in two occasions. We used

both the FOCE-I and the SAEM algorithms to fit the data using models encoded both in ODE and CF solutions. With the selected model, we investigated the effect of the metabolizer status as defined by *CYP* 2D6 and *CYP* 2C19 polymorphisms on the PK of the antipsychotic and its metabolite.

The PK of both compounds were adequately described by a model including a back-transformation mechanism and a first-pass effect with dose fractionation. The acetylation of the metabolite to form back its acetylated parent is a process known for numerous amines [10]. This back-transformation mechanism provided a mechanism-based explanation for the long terminal half-life without accumulation observed for the parent drug. The first-pass effect, on the other hand, allowed to capture the early bump observed in the metabolite data. Auclair et al. also showed that adding a first-pass effect in their model allowed to fit metabolite concentrations that appeared quicker or at the same time than the parent drug [17]. Such a first-pass effect through amidases in the gut and the liver has already been observed in more dramatic proportions for the experimental anti-convulsant related to lidocaïne, D2624 [18]. In order to retain global identifiability of the parameters it was necessary to fix both volumes as equal. We think that this hypothesis is not too strong as both molecules have similar molecular mass and physichochemical properties. In our final model, the rate of appearance of the metabolite, K_{am} is set to the absorption rate of the parent drug K_{ap}. We explained this finding by the absorption rate of the parent drug being rate limiting due to the very quick formation of the metabolite. Duffull et al. also reported this phenomenon for Ivabradine and its metabolite in [19]. The pharmacokinetics of the parent drug and its active metabolite has also been characterized in some phase I studies on healthy volunteers. In one of these studies, concentrations were collected in the same patients following intravenous and oral administrations of the parent drug plus intravenous administration of the metabolite to identify clearances of all processes. The estimates for the percentage of dose that undergoes back-transformation and the fraction of dose that escape first-pass effect obtained in this study were close to the estimates obtained in this population analysis. However, no decrease in bioavailability with the dose was observed in those previous studies.

The same structural model was selected by both NONMEM and MONOLIX whatever the coding approach. Between the two softwares the absorption rate estimates are those that differed the most, but they were also the most difficult to evaluate. With NONMEM, we had trouble obtaining SEs despite using several sets of different initial conditions. On the model where we obtained RSEs, those for the fixed effects ranged from 10 to 58 with NONMEM and from 5 to 25 with MONOLIX. On the other side, NONMEM easily handles steady state concentrations with model encoded in ODE, whereas MLXTRAN requires the user to add dummy lines of dose to mimic the path to steady state. The difference in computing time using this coding approach would thus increase with number of subjects and time to reach steady state. The use of closed form solutions provided an important gain of time for the continuation of the analysis, especially to compute the p-value by permutation. To our knowledge, a comparison of both software has not

yet be performed with such a complex model and using both codings.

CYP 2D6 is an important catalyst of the oxidation of various antipsychotic agents: chlorpromazine, thioridazine, risperidone and haloperidol [20]. Recently, Korean *CYP* 2D6 PM have been shown to have significantly higher risperidone and 9-hydroxyrisperidone (its active metabolite) plasma concentrations [21]. Here, the AUC were higher in *CYP* 2D6 PM patients for both the metabolite and the parent drug. We think that the impact of the *CYP* 2D6 polymorphisms on the parent drug is related to the back-transformation mechanism, indeed Hamrén et al. also described an accumulation of tesaglitar due to the accumulation of its metabolite in patients with impaired renal function [22].

The present work opens some perspectives for the ongoing development of the antipsychotic. The occurrence of *CYP* 2D6 PM is higher in caucasians (5-10%) than in east asians (about 1%) [23] and the *CYP* 2D6 activity is lower in chinese EM compared to caucasians [23] due to the *CYP* 2D6*10 allele. Thus, additionnal investigation should be useful to assess the impact of *CYP* 2D6 polymorphisms in population other than caucasians. From a more methodological perspective, the genetic component of variability R_{GC} was a good predictor of which parameter was under the influence of genetic polymorphisms. This finding should be confirmed after the addition of other covariates that remain unchanged or nearly unchanged over time (such as gender or age).

CONCLUSION

The structural model was rather complex but using the SAEM algorithm we could estimate all the population parameters and the between and within subject variabilities with standard errors. *CYP* 2D6 polymorphisms appeared to impact on the PK of the novel antipsychotic and its metabolite in a large extent.

ACKNOWLEDGMENTS

The authors would like to thank Pr. N. Holford for his clever inputs on the choice of the parameter's notation to ease the understanding of the different structural models. We would also like to thank the MONOLIX development team for making the MLXTRAN tool available for this work.

During this work, Julie Bertrand was supported by a grant from the Institut de Recherches Internationales Servier (France).

References

1. Zhou S, Liu J, Chowbay B. Polymorphism of human cytochrome P450 enzymes and its clinical impact. Drug Metab Rev. 2009;41:89–295.

2. Bertrand J, Comets E, Mentré F. Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. J Biopharm Stat. 2008;18:1084–102.

3. Bertrand J, Comets E, Laffont C, Chenel M, Mentré F. Pharmacogenetics and population pharmacokinetics: impact of the design on three tests using the SAEM algorithm. J Pharmacokinet Pharmacodyn. 2009;36:317–339.

4. Manly B. Randomization, Bootstrap and Monte Carlo Methods in Biology. 2nd ed. London: Chapman & Hall; 1998.

5. Sheiner L, Beal S. NONMEM Version 5.1. San Francisco; 1998.

6. Lavielle M. MONOLIX (MOdèles NOn LInéaires à effets miXtes). Orsay, France; 2008. http://software.monolix.org/index.php.

7. Gibaldi M, Perrier D. Pharmacokinetics. 2nd ed. New York: Marcel Dekker; 1982.

8. Vaughan D, Mallard D, Trainor A, Mitchard M. General Pharmacokinetic Equations for Linear Mammillary Models with Drug Absorption into Peripheral Compartments. Europ J clin Pharmacol. 1975;8:141–8.

9. Schwartz G. Estimating the dimension of a model. Ann Stat. 1978;6:461–464.

10. Cheng H, Jusko W. Pharmacokinetics of reversible metabolic systems. Biopharm Drug Dispos. 1993;14:721–66.

11. Gabrielsson J, Weiner D. Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications. 4th ed. Stockholm: Apotekarsocieteten; 2008.

12. Chou W, Yan F, Robbins-Weilert D, Ryder T, Liu W, Perbost C, et al. Comparison of two CYP2D6 genotyping methods and assessment of genotype-phenotype relationships. Clin Chem. 2003;49:542–51.

13. Ozdemir V, Kalow W, Tang B, Paterson A, Walker S, Endrenyi L, et al. Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration method. Pharmacogenetics. 2000;10:373–88.

14. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2008. http://www.R-project.org. ISBN 3-900051-07-0.

15. Karlsson M, Holford N. A Tutorial on Visual Predictive Checks. 17th Population Approach Group in Europe. 2008;p. Marseille, France Abstr 1434 [www.page–meeting.org/?abstract=1434].

16. Mauri M, Volonteri L, Colasanti A, Fiorentini A, De Gaspari I, Bareggi S. Clinical pharmacokinetics of atypical antipsychotics: a critical review of the relationship between plasma concentrations and clinical response. Clin Pharmacokinet. 2007;46:359–88.

17. Chou W, Yan F, Robbins-Weilert D, Ryder T, Liu W, Perbost C, et al. Comparison of two CYP2D6 genotyping methods and assessment of genotype-phenotype relationships. Clin Chem. 2003;49:542–51.

18. Martin S, Bishop F, Kerr B, Moor M, Moore M, Sheffels P, et al. Pharmacokinetics and metabolism of the novel anticonvulsant agent N-(2,6-dimethylphenyl)-5-methyl-3-isoxazolecarboxamide (D2624) in rats and humans. Drug Metab Dispos. 1997;25:40–6.

19. Duffull S, Chabaud S, Nony P, Laveille C, Girard P, Aarons L. A pharmacokinetic simulation model for ivabradine in healthy volunteers. Eur J Pharm Sci. 2000;10:285–94.

20. Murray M. Role of CYP pharmacogenetics and drug-drug interactions in the efficacy and safety of atypical and other antipsychotic agents. J Pharm Pharmacol. 2006;58:871–85.

21. Kang R, Jung S, Kim K, Lee D, Cho H, Jung B, et al. Effects of CYP2D6 and CYP3A5 genotypes on the plasma concentrations of risperidone and 9-hydroxyrisperidone in Korean schizophrenic patients. J Clin Psychopharmacol. 2009;29:272–7.

22. Hamrén B, Ericsson H, Samuelsson O, Karlsson M. Mechanistic modelling of tesaglitazar pharmacokinetic data in subjects with various degrees of renal function–evidence of interconversion. Br J Clin Pharmacol. 2008;65:855–63.

23. Bertilsson L, Lou Y, Du Y, Liu Y, Kuang T, Liao X, et al. Pronounced differences between native Chinese and Swedish populations in the polymorphic hydroxylations of debrisoquin and S-mephenytoin. Clin Pharmacol Ther. 1992;51:388–97.

LEGEND TO FIGURES

Figure 1. Structural representation of the four structural models regrouped by inclusion of back-transformation mechanism, first-pass effect and dose fractionation.

Figure 2. Observed concentrations of parent drug (a) and metabolite (b) *versus* time collected in two occasions, four and eight weeks after the treatment onset (W4 and W8). Open circles represent observations from patients with a dose of 5 mg, open triangles observations from patients with a dose of 10 mg and crosses observations from patients with a dose of 20 mg.

Figure 3. Visual predictive checks plots of the covariate model. The median and 90^{th} prediction interval from 250 simulated data sets are overlaid to the observed concentrations for the parent drug (top) and the metabolite (bottom) at W4 and W8 for a dose of 5 mg (a. open circles), 10 mg (b. open triangles) and 20 mg (c. crosses).

Figure 4. Box plots of parent drug (a) and metabolite (b) trough concentrations at steady-state simulated from the covariate model for 1000 subjects in each *CYP* 2D6 genotype (PM: poor metabolizer, EM: extensive metabolizer) with a dose of 10 mg. The boxes represent 25th to 75th percentiles, with the 50th percentile shown within the boxes; the 10th and 90th percentiles are shown as capped bars.

Table 1: BIC and fixed effects estimates for the five different structural models using both the FOCE with interaction (FOCE-I) or the SAEM algorithm coded in ordinary differential equations (ODE) or closed form (CF)

	Without interconversion				With interconversion															
	Without first-pass effect								With first-pass effect											
									Without dose fractionation				With dose fractionation							
													Different absorption rates				Similar absorption rates			
	FOCE-I		SAEM		FOCE-I		SAEM		FOCE-I		SAEM		FOCE-I		SAEM		FOCE-I		SAEM	
	ODE	CF	ODE	CF	ODE	CF	ODE	CF	ODE	CF	ODE	CF	ODE	CF	ODE	CF	ODE	CF	ODE	CF
BIC	10432	10441	10427	10417	10317	10326	10384	10330	10232	10240	10234	10201	10194	10200	10206	10158	10184	10194	10165	10136
F_p	-	-	-	-	-	-	-	-	-	-	-	-	0.83	0.83	0.81	0.83	0.83	0.83	0.85	0.86
K_{ap} $(/h)$	9.7	9.9	6.15	9.83	6.9	9.9	5.06	2.1	1.5	1.47	2.57	2.36	3.27	3.24	3.26	5.9	3.01	3.45	2.77	5.51
K_{am} $(/h)$	-	-	-	-	-	-	-	-	0.24	0.25	0.15	0.19	3.7	3.04	4.36	6.44	3.01	3.45	2.77	5.51
V L	25	25.6	25.8	25.6	21	20.8	21.3	19.6	18	17.9	19.4	18.1	18.6	17.6	17.3	17.1	17.7	17.7	17.7	17.4
Cl_{po} (nmol/L)	0.63	0.25	0.21	0.25	1.3	1.32	1.01	1.89	2.1	2.07	1.3	1.66	1.63	1.86	1.56	1.59	1.98	1.85	1.67	1.67
Cl_{pm} $(nmol/L)$	2.3	2.67	2.42	2.37	2.9	2.88	2.79	2.55	1.8	1.74	2.27	1.77	1.89	1.63	1.54	1.56	1.6	1.65	1.82	1.65
Cl_{mo} $(nmol/L)$	0.55	0.62	0.575	0.566	0.41	0.399	0.405	0.267	0.28	0.285	0.379	0.314	0.365	0.313	0.33	0.329	0.29	0.317	0.311	0.314
Cl_{mp} $(nmol/L)$	-	-	-	-	0.19	0.182	0.16	0.226	0.18	0.175	0.159	0.146	0.144	0.146	0.129	0.133	0.16	0.145	0.157	0.147

Table 2: Population pharmacokinetic parameters of parent drug and its active metabolite for the basic and the genetic covariate model: estimates and relative standard errors (RSE)

	Basic model (N=101)		Covariate model (N=99)	
Parameter (unit)	Estimate	RSE (%)	Estimate	RSE (%)
f	1	-	1	-
$\beta_{f,dose}$*	-0.029	19	-0.03	19
F_p	0.84	2	0.86	2
K_a(/h)	6.16	31	7.19	34
VL	18.7	4	18.7	4
Cl_{po} (L/h)	1.32	12	1.35	11
Cl_{pm} (L/h)	2.15	7	2.17	7
$Cl_{mo}L/h$ (L/h)	0.41	9	0.43	9
$\beta_{Cl_{mo},CYP2D6}$**	-	-	-0.629	36
Cl_{mp} (L/h)	0.14	11	0.15	10
ωf (%)	22	22	21	23
ωF_p (%)	0	-	0	-
ωK_a (%)	94	48	93	55
ωV (%)	20	24	19	26
ωCl_{po} (%)	46	33	42	36
ωCl_{pm} (%)	0	-	0	-
ωCl_{mo} (%)	58	13	57	13
ωCl_{mp} (%)	45	72	35	103
γf (%)	15	34	15	34
γF_p (%)	0	-	0	-
γK_a (%)	131	24	137	26
γV (%)	15	34	15	33
γCl_{po} (%)	36	47	37	42
γCl_{pm} (%)	0	-	0	-
γCl_{mo} (%)	29	35	30	32
γCl_{mp} (%)	31	166	29	165
$\sigma Parent$ (%)	28	4	28	4
$\sigma Metabolite$ (%)	9	3	9	3

*$f_i = f \times e^{\beta_{f,dose} \times (DOSE-10)} e^{\eta_{f,i}}$

**$Cl_{mo,i} = Cl_{mo} \times e^{\beta_{Cl_{mo},CYP2D6}} e^{\eta_{Cl_{mo,i}}}$ for PM

Figure 1

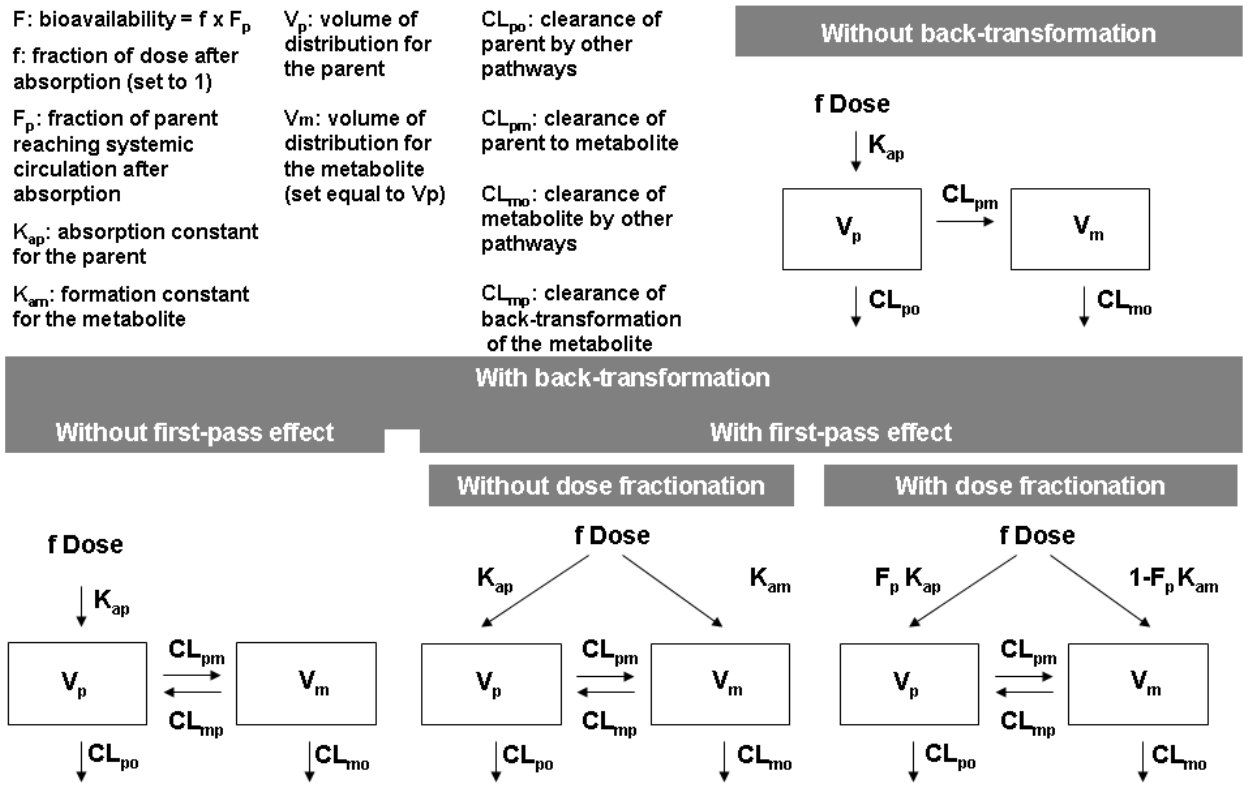

F: bioavailability = f x Fp
f: fraction of dose after absorption (set to 1)
Fp: fraction of parent reaching systemic circulation after absorption
Kap: absorption constant for the parent
Kam: formation constant for the metabolite
Vp: volume of distribution for the parent
Vm: volume of distribution for the metabolite (set equal to Vp)
CLpo: clearance of parent by other pathways
CLpm: clearance of parent to metabolite
CLmo: clearance of metabolite by other pathways
CLmp: clearance of back-transformation of the metabolite
Without back-transformation
f Dose
Kap
Vp
CLpm
Vm
CLpo
CLmo
With back-transformation
Without first-pass effect
With first-pass effect
Without dose fractionation
With dose fractionation
f Dose
Kap
Vp
CLpm
CLmp
Vm
CLpo
CLmo
f Dose
Kap
Kam
Vp
CLpm
CLmp
Vm
CLpo
CLmo
f Dose
Fp Kap
1-Fp Kam
Vp
CLpm
CLmp
Vm
CLpo
CLmo

Figure 2

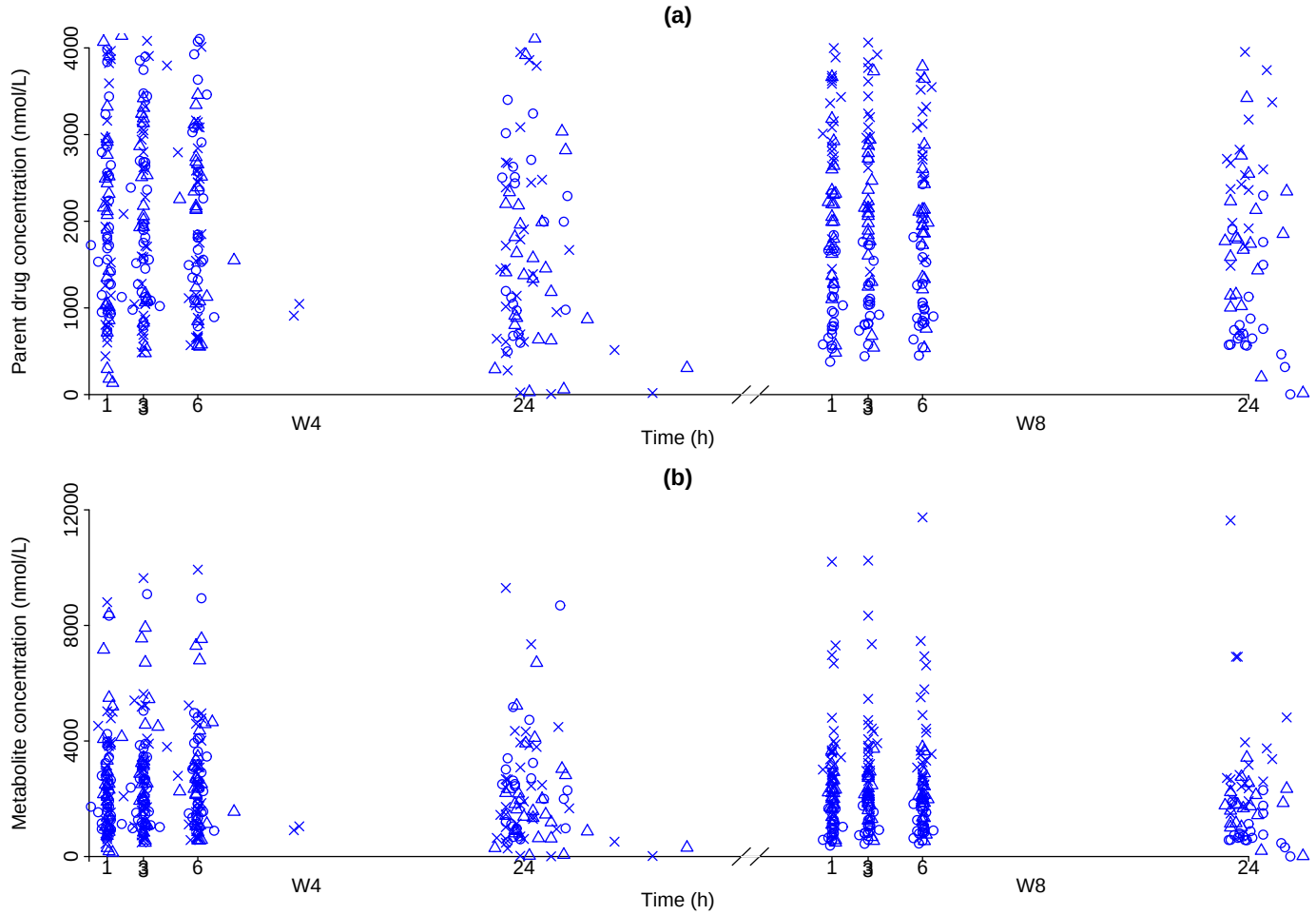

(a)
Parent drug concentration (nmol/L)
0
1000
2000
3000
4000
1
3
6
24
W4
Time (h)
(b)
Metabolite concentration (nmol/L)
0
4000
8000
12000
W8

Figure 3

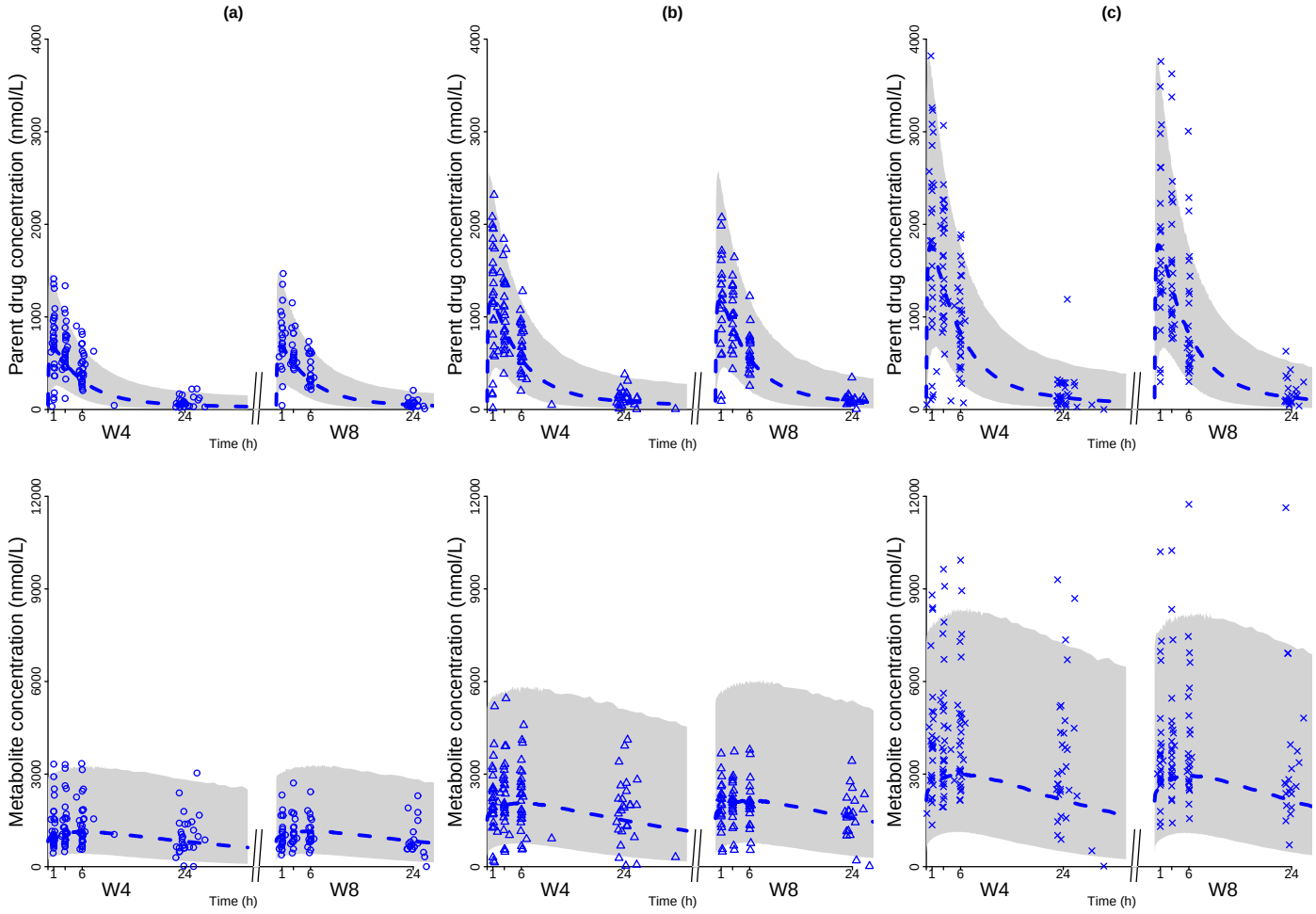
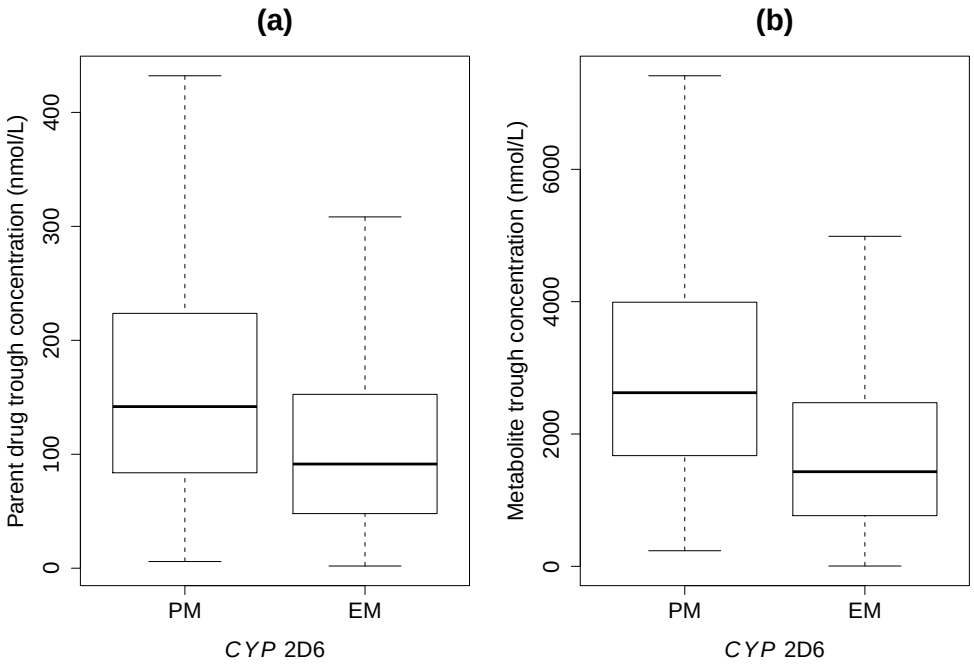

(a)
(b)
(c)
Parent drug concentration (nmol/L)
Metabolite concentration (nmol/L)
0
1000
2000
3000
4000
6000
9000
12000
1
6
24
W4
W8
Time (h)

Figure 4


(a)
Parent drug trough concentration (nmol/L)
0
100
200
300
400
PM
EM
CYP 2D6
(b)
Metabolite trough concentration (nmol/L)
0
2000
4000
6000
PM
EM
CYP 2D6

APPENDIX: Ordinary differentials equations and closed form solutions for the different structural models

- D: dose
- G: pre-systemic compartment
- C_P: parent drug
- C_M: metabolite
- $k_{po} = \frac{CL_{po}}{V}$
- $k_{pm} = \frac{CL_{pm}}{V}$
- $k_{mo} = \frac{CL_{mo}}{V}$
- $k_{mp} = \frac{CL_{mp}}{V}$

Without interconversion

$$\begin{aligned} G(0) =& fD \\ \frac{dG}{dt} =& -K_{ap}G \\ \frac{dC_P}{dt} =& \frac{K_{ap}G}{V} - (k_{po} + k_{pm})C_P \\ \frac{dC_M}{dt} =& -k_{mo}C_M + k_{pm}C_P \end{aligned}$$

- $E_1 = k_{po} + k_{pm}$

$$C_P = \frac{K_{ap}fD}{V}\left[\begin{array}{c} \frac{e^{-K_{ap}t}}{(E_1 - K_{ap})(1 - e^{-K_{ap}\tau})} \\ + \frac{e^{-E_1 t}}{(K_{ap} - E_1)(1 - e^{-E_1\tau})} \end{array}\right]$$

$$C_M = \frac{k_{pm}K_{ap}fD}{V}\left[\begin{array}{c} \frac{e^{-K_{ap}t}}{(E_1 - K_{ap})(k_{mo} - K_{ap})(1 - e^{-K_{ap}\tau})} \\ + \frac{e^{-E_1 t}}{(K_{ap} - E_1)(k_{mo} - E_1)(1 - e^{-E_1\tau})} \\ + \frac{e^{-k_{mo}t}}{(K_{ap} - k_{mo})(E_1 - k_{mo})(1 - e^{-k_{mo}\tau})} \end{array}\right]$$

With interconversion

- $E_1 = k_{po} + k_{pm}$
- $E_2 = k_{mo} + k_{mp}$
- $\lambda_1 = \frac{(E_1+E_2)+\sqrt{(E_1-E_2)^2+4k_{mp}k_{pm}}}{2}$
- $\lambda_2 = \frac{(E_1+E_2)-\sqrt{(E_1-E_2)^2+4k_{mp}k_{pm}}}{2}$

$$
\begin{aligned}
G(0) =& fD \\
\frac{dG}{dt} =& -K_{ap}G \\
\frac{dC_P}{dt} =& \frac{K_{ap}G}{V} - (k_{po}+k_{pm})C_P + k_{mp}C_M \\
\frac{dC_M}{dt} =& -(k_{mo}+k_{mp})C_M + k_{pm}C_P
\end{aligned}
$$

$$
C_P = \frac{K_{ap}fD}{V}\left[\begin{array}{l}
\dfrac{(E_2-K_{ap})e^{-K_{ap}t}}{(\lambda_1-K_{ap})(\lambda_2-K_{ap})(1-e^{-K_{ap}\tau})} \\
+\dfrac{(E_2-\lambda_1)e^{-\lambda_1 t}}{(K_{ap}-\lambda_1)(\lambda_2-\lambda_1)(1-e^{-\lambda_1\tau})} \\
+\dfrac{(E_2-\lambda_2)e^{-\lambda_2 t}}{(K_{ap}-\lambda_2)(\lambda_1-\lambda_2)(1-e^{-\lambda_2\tau})}
\end{array}\right]
$$

$$
C_M = \frac{k_{pm}K_{ap}fD}{V}\left[\begin{array}{l}
\dfrac{e^{-K_{ap}t}}{(\lambda_1-K_{ap})(\lambda_2-K_{ap})(1-e^{-K_{ap}\tau})} \\
+\dfrac{e^{-\lambda_1 t}}{(K_{ap}-\lambda_1)(\lambda_2-\lambda_1)(1-e^{-\lambda_1\tau})} \\
+\dfrac{e^{-\lambda_2 t}}{(K_{ap}-\lambda_2)(\lambda_1-\lambda_2)(1-e^{-\lambda_2\tau})}
\end{array}\right]
$$

With interconversion and first-pass effect

$$
\begin{aligned}
G(0) =& fD \\
\frac{dG}{dt} =& -(K_{ap}+K_{am})G \\
\frac{dC_P}{dt} =& \frac{K_aG}{V} - (k_{po}+k_{pm})C_P \\
& + k_{mp}C_M \\
\frac{dC_M}{dt} =& \frac{K_{am}G}{V} - (k_{mo}+k_{mp})C_M \\
& + k_{pm}C_P
\end{aligned}
$$

$$C_P = \frac{K_{ap} f D}{V} \left[\begin{array}{l} \frac{(E_2 - (K_{ap} + K_{am}))e^{-(K_{ap}+K_{am})t}}{(\lambda_1 - (K_{ap} + K_{am}))(\lambda_2 - (K_{ap} + K_{am}))} \\ + \frac{(E_2 - \lambda_1)e^{-\lambda_1 t}}{((K_{ap} + K_{am}) - \lambda_1)(\lambda_2 - \lambda_1)} \\ + \frac{(E_2 - \lambda_2)e^{-\lambda_2 t}}{((K_{ap} + K_{am}) - \lambda_2)(\lambda_1 - \lambda_2)} \end{array} \right]$$

$$+ \frac{k_{mp} K_{am} f D}{V} \left[\begin{array}{l} \frac{e^{-(K_{ap}+K_{am})t}}{(\lambda_1 - (K_{ap} + K_{am}))(\lambda_2 - (K_{ap} + K_{am}))} \\ + \frac{e^{-\lambda_1 t}}{((K_{ap} + K_{am}) - \lambda_1)(\lambda_2 - \lambda_1)} \\ + \frac{e^{-\lambda_2 t}}{((K_{ap} + K_{am}) - \lambda_2)(\lambda_1 - \lambda_2)} \end{array} \right]$$

$$C_M = \frac{k_{pm} k_{mp} K_{am} f D}{V} \left[\begin{array}{l} \frac{e^{-E_2 t}}{((K_{ap} + K_{am}) - E_2)(\lambda_1 - E_2)(\lambda_2 - E_2)} \\ + \frac{e^{-(K_{ap}+K_{am})t}}{(E_2 - (K_{ap} + K_{am}))(\lambda_1 - (K_{ap} + K_{am}))(\lambda_2 - (K_{ap} + K_{am}))} \\ + \frac{e^{-\lambda_1 t}}{(E_2 - \lambda_1)((K_{ap} + K_{am}) - \lambda_1)(\lambda_2 - \lambda_1)} \\ + \frac{e^{-\lambda_2 t}}{(E_2 - \lambda_2)((K_{ap} + K_{am}) - \lambda_2)(\lambda_1 - \lambda_2)} \end{array} \right]$$

$$+ \frac{K_{am} f D}{V} \left[\frac{e^{-E_2 t}}{(K_{ap} + K_{am}) - E_2} + \frac{e^{-(K_{ap}+K_{am})t}}{E_2 - (K_{ap} + K_{am})} \right]$$

$$+ \frac{k_{pm} K_{ap} f D}{V} \left[\begin{array}{l} \frac{e^{-(K_{ap}+K_{am})t}}{(\lambda_1 - (K_{ap} + K_{am}))(\lambda_2 - (K_{ap} + K_{am}))} \\ + \frac{e^{-\lambda_1 t}}{((K_{ap} + K_{am}) - \lambda_1)(\lambda_2 - \lambda_1)} \\ + \frac{e^{-\lambda_2 t}}{((K_{ap} + K_{am}) - \lambda_2)(\lambda_1 - \lambda_2)} \end{array} \right]$$

With interconversion, first-pass effect and dose fractionment

$$\begin{aligned} G_1(0) &= F_p f D \\ G_2(0) &= (1 - F_p) f D \\ \frac{dG_1}{dt} &= - K_{ap} G_1 \\ \frac{dG_2}{dt} &= - K_{am} G_2 \\ \frac{dC_P}{dt} &= \frac{K_{ap} G_1}{V} - (k_{po} + k_{pm}) C_P + k_{mp} C_M \\ \frac{dC_M}{dt} &= \frac{K_{am} G_2}{V} - (k_{mo} + k_{mp}) C_M + k_{pm} C_P \end{aligned}$$

$$C_P = \frac{K_{ap}F_pfD}{V}\left[\begin{array}{c}\frac{(E_2-K_{ap})e^{-K_{ap}t}}{(\lambda_1-K_{ap})(\lambda_2-K_{ap})(1-e^{-K_{ap}\tau})}\\ +\frac{(E_2-\lambda_1)e^{-\lambda_1 t}}{(K_{ap}-\lambda_1)(\lambda_2-\lambda_1)(1-e^{-\lambda_1\tau})}\\ +\frac{(E_2-\lambda_2)e^{-\lambda_2 t}}{(K_{ap}-\lambda_2)(\lambda_1-\lambda_2)(1-e^{-\lambda_2\tau})}\end{array}\right]$$

$$+\frac{k_{mp}K_{am}(1-F_p)fD}{V}\left[\begin{array}{c}\frac{e^{-K_{am}t}}{(\lambda_1-K_{am})(\lambda_2-K_{am})(1-e^{-K_{am}\tau})}\\ +\frac{e^{-\lambda_1 t}}{(K_{am}-\lambda_1)(\lambda_2-\lambda_1)(1-e^{-\lambda_1\tau})}\\ +\frac{e^{-\lambda_2 t}}{(K_{am}-\lambda_2)(\lambda_1-\lambda_2)(1-e^{-\lambda_2\tau})}\end{array}\right]$$

$$C_M = \frac{k_{pm}k_{mp}K_{am}(1-F_p)fD}{V}\left[\begin{array}{c}\frac{e^{-E_2 t}}{(K_{am}-E_2)(\lambda_1-E_2)(\lambda_2-E_2)(1-e^{-E_2\tau})}\\ +\frac{e^{-K_{am}t}}{(E_2-K_{am})(\lambda_1-K_{am})(\lambda_2-K_{am})(1-e^{-K_{am}\tau})}\\ +\frac{e^{-\lambda_1 t}}{(E_2-\lambda_1)(K_{am}-\lambda_1)(\lambda_2-\lambda_1)(1-e^{-\lambda_1\tau})}\\ +\frac{e^{-\lambda_2 t}}{(E_2-\lambda_2)(K_{am}-\lambda_2)(\lambda_1-\lambda_2)(1-e^{-\lambda_2\tau})}\end{array}\right]$$

$$+\frac{K_{am}(1-F_p)fD}{V}\left[\begin{array}{c}\frac{e^{-E_2 t}}{(K_{am}-E_2)(1-e^{-E_2\tau})}\\ +\frac{e^{-K_{am}t}}{(E_2-K_{am})(1-e^{-K_{am}\tau})}\end{array}\right]$$

$$+\frac{k_{pm}K_{ap}F_pfD}{V}\left[\begin{array}{c}\frac{e^{-K_{ap}t}}{(\lambda_1-K_{ap})(\lambda_2-K_{ap})(1-e^{-K_{ap}\tau})}\\ +\frac{e^{-\lambda_1 t}}{(K_{ap}-\lambda_1)(\lambda_2-\lambda_1)(1-e^{-\lambda_1\tau})}\\ +\frac{e^{-\lambda_2 t}}{(K_{ap}-\lambda_2)(\lambda_1-\lambda_2)(1-e^{-\lambda_2\tau})}\end{array}\right]$$

3.3 Effet du polymorphisme *CYP*2B6 G516T sur pharmacocinétique de la névirapine chez des patients cambodgiens infectés par le VIH

3.3.1 Introduction

La névirapine est une dipyridodiazépine agissant comme un inhibiteur non compétitif de la transcriptase inverse produite par le rétrovirus d'immunodéficience humaine lorsqu'il a pénétré sa cellule cible : les cellules du système immunitaire. Dans les pays en voie de développement, où les ressources thérapeutiques sont plus limitées, la névirapine associée à deux analogues nucléosidiques de la transcriptase inverse est la ligne thérapeutique recommandée pour les patients naïfs de traitement. Au Cambodge, la prévalence du syndrome d'immunodéficience acquise humaine (SIDA) dans la population générale (âgée de 15 à 49 ans) est passé de 2% à 0.6% entre 1998 et 2006. Cette baisse à eu lieu suite à un grand nombre de décès dans les premières années de l'épidémie qui se sont produits avant l'installation en 2001 d'un système de soin géré par le *National Centre for HIV/AIDS Dermatology and STDs* (NCHADS) qui supervise la prévention, les actes médicaux et les traitements. A la fin de l'année 2008, 29000 patients étaient sous traitement par antirétroviraux et 69.5% d'entre eux recevaient de la névirapine.

La pharmacocinétique de la névirapine est caractérisée par une longue demi-vie avec une forte propension à se lier aux protéines plasmatiques (60%) et une élimination principalement métabolique par les enzymes CYP3A et CYP2B6 (Riska et al., 1999; Erickson et al., 1999). A ce jour, une seule étude a été réalisée chez des patients cambodgiens sur l'efficacité de la névirapine (Ferradini et al., 2007), et aucune étude pharmacocinétique. Le polymorphisme *CYP* 2B6 G516T a été associé à une diminution de l'expression de l'enzyme dans le foie et Rotger et al. (2005) ainsi que Penzak et al. (2007) ont montré que ce polymorphisme était associé à des concentrations élevées de névirapine. Or les fréquences alléliques des polymorphismes génétiques codant pour les enzymes métaboliques ont été peu renseignées dans les populations asiatiques, et des fréquences déterminées à partir de données recueillies, par exemple dans la population chinoise, ne peuvent pas forcément être extrapolées à la population cambodgienne.

Dans ce contexte, l'objectif de l'essai PECAN-ANRS 12114 était de caractériser la pharmacocinétique de la névirapine dans la population cambodgienne et d'identifier les facteurs de variabilité.

3.3.2 Méthodes

Présentation de l'essai

Les patients inclus dans l'essai ouvert et monocentrique PECAN-ANRS 12114 proviennent de la cohorte ESTHER réalisée à l'hôpital Calmette à Phnom Penh au

Cambodge. Cette cohorte a débuté en 2003, lorsque les patients atteints du SIDA ont pu avoir accès à des traitements au Cambodge. Pendant les 2 premières semaines de traitement, la névirapine était administrée en une prise journalière de 200 mg, puis pendant le reste du traitement en deux prises de 200 mg en association avec deux prises de 30 mg de stavudine et deux prises de 150 mg de lamivudine. Après six mois de traitement, la stavudine était remplacée par deux prises de 300 mg de zidovudine chez la plupart des patients. Les patients venaient à la clinique tous les mois pour renouveler leurs médicaments mais les charges virales n'étaient pas mesurées à ces occasions.

Tous les six mois, les marqueurs de l'activité hépatique (ALAT, ASAT,γGT, bilirubine), rénale (créatinine, clairance à la créatinine) et immunitaire (niveau des CD4 plasmatiques) étaient renseignés. En deux occasions après 18 et 36 mois de traitement (M18 et M36), les charges virales et les concentrations résiduelles de névirapine avant la prise du matin ont été recueillies en plus de ces marqueurs, et l'observance était recueillie sous la forme d'une échelle d'évaluation analogique. La consommation de tabac et d'alcool était renseignée par un questionnaire. Certains patients étaient testés pour une infection par les hépatites virales C et B. A la deuxième occasion, un prélèvement sanguin était réalisé chez les patients ayant signé un consentement pour effectuer les génotypages.

Une pharmacocinétique plus complète était réalisée chez 10 patients. Ces patients étaient à jeun avant la prise de leur traitement. Les prélèvements ont été recueillis avant la prise de la dose puis 1h, 2h, 4h et 8h après administration.

La modélisation pharmacocinétique

La modélisation des profils de concentration de la névirapine a été réalisée avec le logiciel MONOLIX version 2.4 (Lavielle, 2008). Les concentrations ont été décrites par un modèle à un compartiment d'absorption et d'élimination d'ordre 1 à l'équilibre, paramétré en volume (V/F) et clairance (CL/F) apparents. Les données sous 0.05 ng/mL (la limite de quantification ou LOQ) étaient retirées de l'analyse.

Dans un premier temps, la matrice de variance-covariance des effets aléatoires ainsi que le modèle de l'erreur résiduelle ont été déterminé sur les concentrations recueillies dans l'étude pharmacocinétique complète et sur les concentrations résiduelles obtenues à M36. Le choix du modèle d'erreur (additif, proportionnel ou combiné) et des variances inter-sujets (IIV) estimées était basé sur le critère d'information de Bayes (BIC, Schwartz (1978)). Dans un deuxième temps, les concentrations recueillies à M18 ont été ajoutées au jeu de données et des variances inter-occasions (IOV) ont été estimées sur les paramètres ayant une variance inter-sujets non nulle. Les effets aléatoires inter-sujets et occasions ont été intégrés dans l'analyse par un modèle additif sur les log-paramètres.

Analyse des covariables

Les covariables continues considérées étaient l'âge, le poids, la créatinine, la clairance à la créatinine, la charge virale, le taux de CD4 et l'observance, et pour les covariables catégorielles le sexe, les co-traitements, avoir une charge virale au dessus de 400 copies/mL, le statut co-infecté par le virus de l'hépatite B ou C, et les génotypes pour les polymorphismes *CYP* 3A5 A6986G, 2B6 C1459T, 2B6 G516T et ABCB1 C3534T. Nous avons considéré le modèle de covariables suivant pour le paramètre θ du sujet i à l'occasion k avec un effet additif de la covariable sur le logarithme du paramètre, tel que :

$$log(\theta_{i,k}) = log(\theta) + \beta_1 \times cov1_i + \beta_2 \times cov2_{i,k}$$

Où β_1 est le coefficient d'effet associé à une covariable n'évoluant pas au cours du temps (comme le génotype ou le sexe) et β_2 le coefficient d'effet associé à une covariable qui change d'une occasion sur l'autre. Pour faciliter l'interprétation des résultats, les covariables continues ont été log-transformées et centrées sur le logarithme de la médiane. Afin de déterminer dans quelle mesure les paramètres du modèle étaient susceptibles d'être influencés par les polymorphismes génétiques, nous avons calculé le coefficient de variabilité génétique décrit par Ozdemir et al. (2000) : $R_{GC} = 1 - \frac{IOV}{IIV}$ qui est d'autant plus proche de 1 que le paramètre est susceptible d'être influencé par la variabilité génétique.

La construction du modèle de covariables a été faite par sélection ascendante, l'inclusion dans le modèle étant déterminée par un test de Wald. Lors de la première étape, c'est-à-dire pour les analyses univariées, aucune imputation des données manquantes n'a été réalisée et le seuil de sélection était de 10%. Après l'inclusion de la première covariable dans le modèle, le seuil du test a été ramené à 5% et les données manquantes des autres covariables sélectionnées ont été imputées à la valeur de l'occasion la plus proche si renseignée, et à la médiane sinon.

Une approche de permutation a été réalisée pour évaluer les p-values des covariables conservées dans le modèle final. La permutation à été réalisée sur l'ensemble des patients ayant une ou plusieurs occasions. Lorsque le nombre d'occasions renseignées pour la covariable excédait le nombre d'occasions renseignées pour la réponse, la covariable de l'occasion la plus proche était utilisée. Lorsque le nombre d'occasions renseignées pour la réponse excédait le nombre d'occasions renseignées pour la covariable, la covariable de l'occasion la plus proche était répétée.

Deux cent cinquante jeux de données ont été simulés avec les paramètres estimés par le modèle final. Les vecteurs des $5^{ièmes}$, $50^{ièmes}$ et $95^{ièmes}$ percentiles récupérés sur les concentrations prédites pour chaque temps ont été superposés aux données sur un graphique de type *visual predictive check* (Karlsson et Holford, 2008). De plus, les concentrations observées ont été comparées graphiquement aux concentrations prédites

avec les paramètres de population et avec les paramètres individuels.

3.3.3 Résultats

Les caractéristiques des patients

Cent soixante-dix (170) patients de la cohorte ESTHER ont été inclus dans l'essai PECAN. Cent quarante-cinq (145) patients (80 hommes et 65 femmes) ont participé à l'évaluation à M18 et 161 patients (89 hommes et 72 femmes) à l'évaluation à M36. Dix (10) patients ont participé à l'étude pharmacocinétique complète. L'ensemble de caractéristiques des 170 patients de l'essai à M36 (ou M18 si manquant à M36) est résumé dans la table 3.1.

Tab. 3.1 – Caractéristiques des 170 patients à M36

Covariable	Médiane [range]	N
Âge (an)	36.5 [21.0 ; 64.0]	170
Poids (kg)	55.0 [35.5 ; 82.0]	169
ALAT (UI/mL)	28.5 [11.0 ; 212.0]	166
ASAT (UI/mL)	30.0 [17.0 ; 117.0]	166
γGT (UI/mL)	65 [15.0 ; 687.0]	165
Bilirubine (μmol/L)	7.0 [5.0 ; 37.0]	166
Créatinine (μmol/L)	80.5 [44.0 ; 136.0]	166
Clairance à la créatinine (mL/min)	82.5 [44.3 ; 144.2]	165
CD4 (cellules/mL)	295.0 [14.0 ; 1054.0]	167
Charge virale (copies/mL)	400.0 [20 ; 190530.0]	166
Observance*	10.0 [7.0 ; 10.0]	167
	Effectif (%)	N
Sexe (F/H)	75 (44.1) / 95 (55.9)	170
Patients avec charge virale<400 copies/mL (o/n)	155 (93.4) / 11 (6.6)	166
Patients co-infectés VHC (o/n)	11 (7.6) / 133 (92.4)	144
Patients co-infectés VHB (o/n)	20 (13.7) / 126 (86.3)	146

*L'observance était recueillie sous la forme d'une échelle d'évaluation analogique cotée de 0 à 10

La plupart des patients (86%) étaient sous névirapine + lamivudine/stavudine à M18 et sous névirapine + lamivudine/zidovudine à M36. A M18, 2 patients étaient sous rifampicine et 25 sous fluconazole. A M36, l'âge médian de la population était 37.0 [21.0 ; 64.0] ans et le poids médian 55.0 [36.0 ; 82.0] kg. Seuls 4 patients ont déclaré une consommation régulière de tabac et 1 patient a déclaré une consommation quotidienne d'alcool. Trente-sept pour cent (37%) des patients avaient au moins une constante métabolique (ALAT, ASAT, γGT et bilirubine) deux fois au dessus de la normale. Les patients co-infectés par le VHB ou le VHC représentaient 13.7 et 7.6% des 146 et 144

patients testés ; seuls deux patients étaient co-infectés VHC et VHB. A M18, 88% des patients avaient une charge virale en dessous de 400 copies/mL, ces patients étaient donc maintenus sous névirapine. Quatre-vingt-onze pour cent (91%) d'entre eux avait une charge virale en dessous de ce seuil à M36. L'indice d'observance était particulièrement haut dans cette population où 63% et 57% des patients avaient une observance de 10 à M18 puis M36.

Les polymorphismes génétiques

Les fréquences des polymorphismes *CYP* 3A5 A6986G, *CYP* 2B6 G516T et C1459T et ABCB1 C3435T sont présentées dans la table 3.2. Les proportions observées des génotypes ne différaient pas significativement des proportions d'Hardy-Weinberg (Crow, 1999) pour tous les polymorphismes à l'étude. Aucun patient n'était homozygote rare pour le polymorphisme *CYP* 2B6 C1459T.

Tab. 3.2 – Fréquences génotypiques et alléliques chez les 170 patients

Polymorphisme génétique	Génotype	N	Allèle	Fréquence	P-value*
CYP 3A5 A6986G	AA	20	A	0.35	1
	AG	75	G	0.65	
	GG	70			
	NR	5			
CYP 2B6 G516T	GG	61	G	0.65	0.49
	GT	78	T	0.35	
	TT	14			
	NR	17			
CYP 2B6 C1459T	CC	155	C	0.99	0.98
	CT	4	T	0.01	
	NR	11			
ABCB1 C3435T	CC	69	C	0.62	0.36
	CT	68	T	0.38	
	TT	30			
	NR	3			

NR : non renseigné

* Test des proportions d'Hardy-Weinberg

Les concentrations de névirapine

Au total, 316 concentrations ont été recueillies chez 169 patients. Quatre patients avaient des concentrations aux 3 occasions, 136 patients à 2 occasions, et 29 patients à 1

seule occasion. Un patient a été retiré de l'analyse car sa seule concentration à M18 était sous la LOQ, trois autres concentrations étaient sous la LOQ, 2 à M18 et 1 à M36. Ce faible nombre de concentrations indétectables est en accord avec la bonne observance rapportée par l'ensemble des patients. La figure 3.1 représente les concentrations de névirapine recueillies aux 3 occasions, M18, M36 et lors de l'étude pharmacocinétique complète. La médiane des concentrations résiduelles était de 5705 [0.05 ; 13871] ng/mL et 5709 [0.05 ; 15422] ng/mL à M18 et M36. Il est à noter que 17.2 et 20.5% des patients à M18 et M36 avaient des concentrations de névirapine supérieure à 8000 ng/mL, qui est la marge supérieure de la zone de concentrations "efficaces" admises pour la névirapine (Yeni et groupe des experts "Prise en charge médicale des personnes infectées par le VIH", 2006).

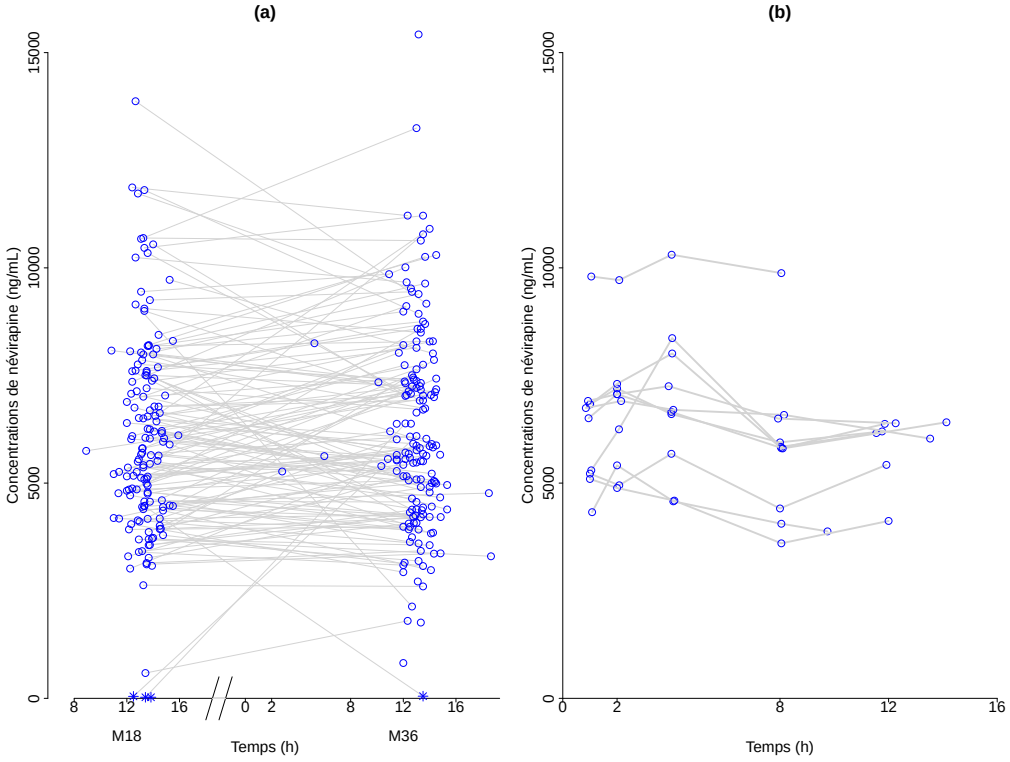


Fig. 3.1 – Concentrations plasmatiques de névirapine observées en fonction du temps chez 170 patients cambodgiens atteints du SIDA recueillis lors des évaluations à 18 et 36 mois de traitement (a) et lors de l'étude pharmacocinétique complète (b). Les concentrations d'un même sujet sont reliées par un trait gris.

La pharmacocinétique de la névirapine

Les concentrations de névirapine ont été adéquatement décrites par un modèle à un compartiment d'absorption et d'élimination d'ordre 1. Avec le modèle basique, la clairance apparente de la névirapine était estimée à 2.67 L/h avec une variabilité inter-sujets de 28.5% et une variabilité inter-occasions de 17.1%. La constante d'absorption et le volume apparent estimés étaient de 1.73 /h et 213 L. Estimer les variances inter-sujets

sur ces paramètres n'améliorait pas le modèle en terme de BIC. Ces variances n'étaient probablement pas identifiables sur ce plan d'expérience où seule une concentration résiduelle était recueillie par occasion chez la plupart des patients. Un modèle d'erreur additif a été sélectionné avec un talon estimé à 519 ng/mL. Les estimations des paramètres du modèle basique et leur erreur d'estimation relative (%) sont données dans la table 3.3. Le coefficient R_{GC} de la clairance de la névirapine était de 36%.

TAB. 3.3 – Estimations des paramètres et de leur erreur d'estimation relative (EER) pour le modèle basique (N=169) et le modèle final (N=152)

	Modèle basique		Modèle final		
Paramètre	Estimation	EER (%)	Estimation	EER (%)	P-value*
k_a /h	1.64	85	1.58	92	
V/F L	213	33	223	35	
CL/F L/h	2.67	3	2.95	4	
$\beta_{CYP\ 2B6\ GT}$			-0.12	41	0.01
$\beta_{CYP\ 2B6\ TT}$			-0.46	18	$9.9.10^{-4}$
$\beta_{clairance\ créatinine}$			0.23	37	$6.9.10^{-3}$
$\omega_{CL/F}$ (%)	28	8	24	8	
$\lambda_{CL/F}$ (%)	17	8	17	9	
$\sigma(ng/mL)$	520	12	580	11	

ω = écart type de la variabilité inter-sujets
λ = écart type de la variabilité inter-occasions
* Test de Wald par permutation

Après la première étape de sélection, le polymorphisme *CYP* 2B6 G516T (p-value = 0.02 et 3.10^{-10} respectivement pour les génotypes GT et TT), la clairance à la créatinine (p-value = 0.07) et le statut co-infecté par le VHC (p-value = 0.04) étaient significativement associés à la clairance de la névirapine au seuil de 10%. Après la sélection ascendante par le test de Wald, le modèle conservait uniquement l'effet du polymorphisme *CYP* 2B6 G516T (p-value = 0.01 et $4.2.10^{-8}$) et de la clairance à la créatinine (p-value = $6.7.10^{-3}$). Les p-values obtenues par permutation pour ces covariables étaient respectivement 0.01, $9.9.10^{-4}$ et $6.9.10^{-3}$. Les estimations des paramètres du modèle final et leur erreur d'estimation relative (%) sont données dans la table 3.3.

La fig 3.2 regroupe les graphiques de validation du modèle, avec les graphiques de type *visual predictive check* et les valeurs observées en fonction des valeurs prédites par les paramètres de population et les paramètres individuels. Ces graphiques sont représentés pour chacun des génotypes du polymorphisme *CYP* 2B6 G516T, la clairance à la créatinine expliquant une part plus négligeable de la variabilité.

La figure 3.3 illustre les effets du polymorphisme *CYP* 2B6 G516T et de la clairance à la créatinine sur la clairance apparente de la névirapine. Cette dernière était estimée à

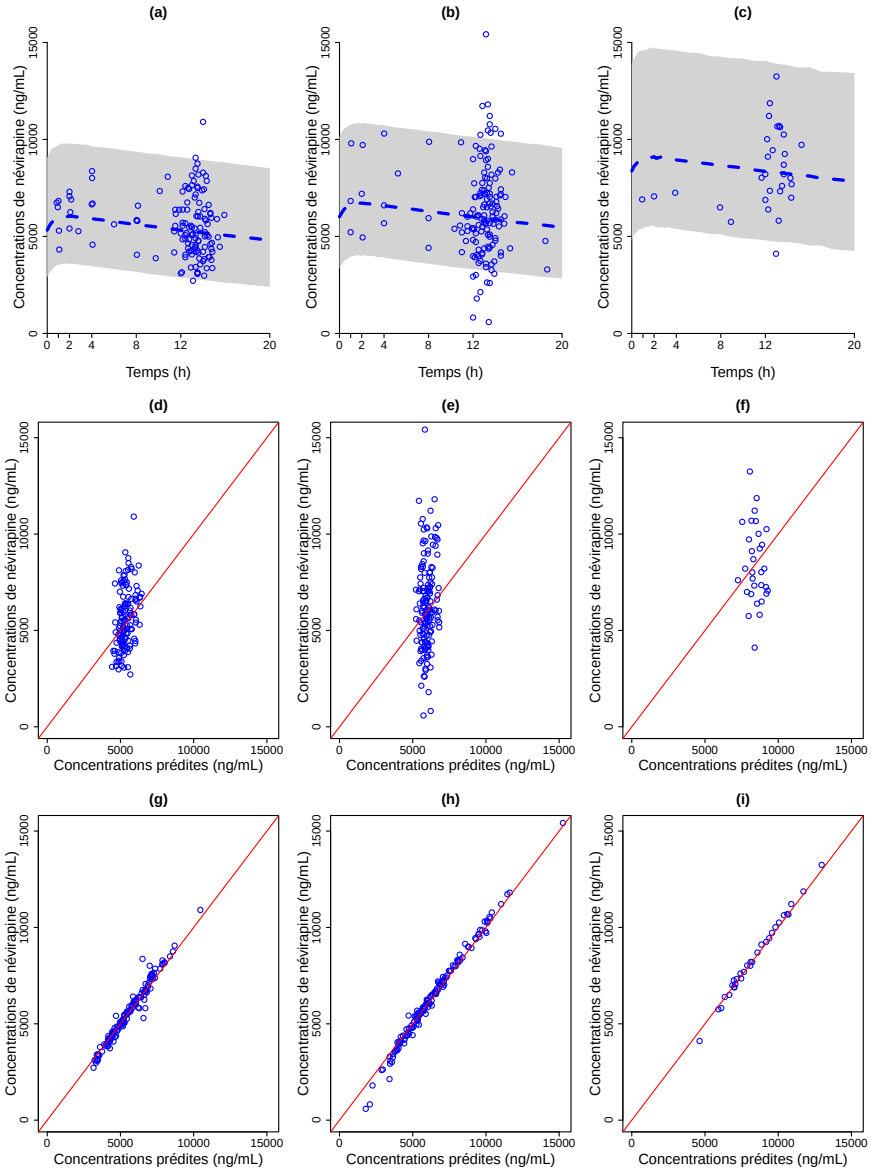


FIG. 3.2 – Concentrations de névirapine en fonction du temps aux 3 occasions superposées à l'intervalle à 90% et à la médiane des prédictions du modèle final regroupées par génotype pour le polymorphisme *CYP* 2B6 G516T : GG (a), GT (b) et TT (c). Concentrations de névirapine observées en fonction des concentrations prédites par les paramètres de population (d, e et f) et par les paramètres individuels (g, h et i) du modèle final regroupées par génotype pour le polymorphisme *CYP* 2B6 G516T. Sur ces 6 graphiques, la figure est divisée en deux par la première bissectrice.

2.95 L/h, 2.62 L/h (p-value = 0.01) et 1.86 L/h (p-value = $9.9.10^{-4}$) pour les patients GG, GT et TT pour le polymorphisme *CYP* 2B6. La plus faible valeur observée de clairance à la créatinine était associée à une diminution de 14% de CL/F alors que la valeur la plus élevée de clairance à la créatinine était associée à une augmentation de 13% de CL/F. L'ajout du polymorphisme *CYP* 2B6 G516T dans le modèle diminuait la variabilité inter-sujets de 2.4%.

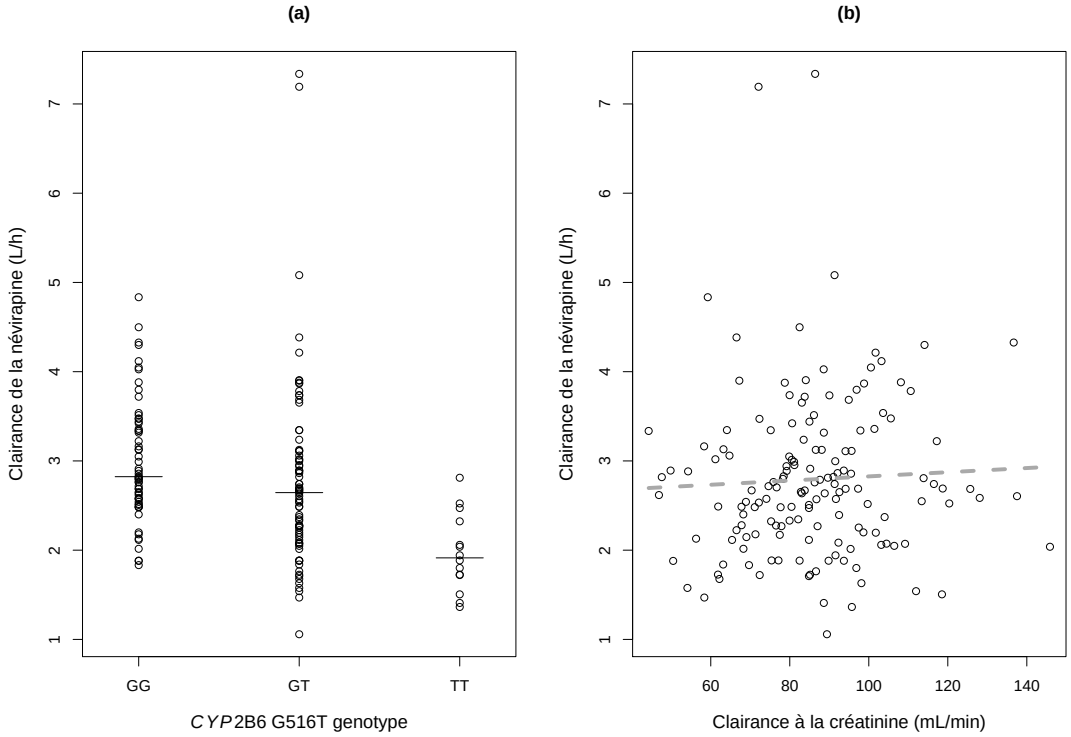


FIG. 3.3 – Estimations bayésiennes empiriques moyennes des clairances apparentes individuelles de la névirapine aux différentes évaluations regroupées par génotype pour le polymorphisme *CYP* 2B6 G516T avec les médianes estimées dans chaque groupe (a) et représentées en fonction des mesures individuelles moyennes de clairance à la créatinine aux différentes évaluations (b). La ligne de tirets grise représente la droite de régression.

3.3.4 Discussion

L'étude par une approche de population de la pharmacocinétique de la névirapine et l'identification des facteurs expliquant sa variabilité ont été réalisées sur un groupe homogène de patients cambodgiens atteints du SIDA particulièrement observants. En effet, l'efficacité du traitement par la névirapine dans cet essai est impressionnante : 90% des patients en succès virologique à M18 l'étaient encore à M36. Une telle observance a déjà été rapportée dans d'autres cohortes cambodgiennes. Entre autres SPIRE et al. (2008) soulignent dans leur étude que 95% des patients étaient observants après 24 mois de traitement. Les traitements antirétroviraux ainsi que des programmes éducationnels sont délivrés gratuitement au Cambodge dans le cadre du NCHADS.

Les paramètres pharmacocinétique obtenus dans cette analyse étaient dans la fourchette des valeurs rapportées dans les autres études pour la constante d'absorption [1.04 ; 1.8] /h (MOLTÓ et al., 2008; DAILLY et al., 2004) et la clairance apparente [1.09 ; 3.3] L/h (DAILLY et al., 2004; HAAS et al., 2009). Par contre, le volume de distribution apparent estimé était légèrement au dessus de la fourchette des valeurs trouvées dans la littérature [75 ; 208] L. Les variabilités inter-sujets et occasions sur la clairance apparente

de la névirapine étaient respectivement en dessous de 30% et 20% avec des erreurs d'estimations de moins de 10%.

Cette étude a permis de renseigner les fréquences des polymorphismes *CYP* 3A5 A6986G, 2B6 C1459T, 2B6 G516T et ABCB1 C3534T dans la population cambodgienne. La fréquence de l'allèle *CYP* 3A5*3 était de 0.65 dans la population de l'essai, soit assez proche de celle estimée dans la population malaisienne (0.6, Balram et al. (2003)). Cette fréquence était plus élevée que celles rapportées dans la population afro-américaine [0.29 ; 0.40] (Haas et al., 2009; Blanco et al., 2002) et plus faible que celles observées dans la population caucasienne [0.8 ; 0.88] (Blanco et al., 2002; Arab-Alameddine et al., 2009) ou chinoise 0.76 (Balram et al., 2003). La fréquence estimée du polymorphisme *CYP* 2B6 C1459T était particulièrement faible 0.01, ce qui est le cas dans la plupart des populations humaines (International HapMap Consortium, 2003). Pour le polymorphisme *CYP* 2B6 G516T, la fréquence estimée de 0.35 était proche de la fréquence rapportée dans les populations chinoises (0.4, Tong et al. (2006)) et afro-américaines (0.38, Haas et al. (2009)), et au dessus de celle rapportée dans la population caucasienne (0.26, Haas et al. (2004)). Enfin, la fréquence estimée de 0.38 pour l'allèle ABCB1 C3534T était comprise dans la fourchette des fréquences alléliques rapportées pour les populations japonaises [0.38 ; 0.49] et malaisiennes [0.37 ; 0.52], légèrement en dessous de celles définies pour les populations caucasiennes [0.46 ; 0.59] et chinoises [0.4 ; 0.53], toutes au dessus de la fourchette observée dans la population africaine [0.1 ; 0.26] (Marzolini et al., 2003).

Dans cette étude nous avons montré un effet du polymorphisme *CYP* 2B6 G516T sur la clairance apparente de la névirapine. Cet effet résulte notamment en une demi-vie estimée de 58.4 h (+/- 17.4 h), 59.0 h (+/- 19.7 h) et 82.9 h (+/- 27.6 h) respectivement chez les patients GG, GT et TT pour ce polymorphisme. Dans leur étude, Haas et al. (2009) ne trouvaient pas d'association significative du polymorphisme *CYP* 2B6 G516T avec la pharmacocinétique de la névirapine après une dose unique. Mais le métabolisme de la névirapine connaissant d'importants phénomènes d'auto-induction (Cheeseman et al., 1995; Riska et al., 1999), les différences observées à l'équilibre peuvent ne pas être détectables après une dose unique. La diminution graduelle de la clairance apparente entre les génotypes ne correspond pas aux modèles génétiques de type dominant ou récessif et pourrait indiquer un effet haplotypique. Toutefois, nous n'avons pas mis en évidence d'effet du polymorphisme *CYP* 2B6 C1459T sur la pharmacocinétique de la névirapine dont l'allèle T est très rare. Une perspective de notre travail serait de considérer un modèle additif ou une approche comme celle proposée par Arab-Alameddine et al. (2009) qui permet de résumer d'une part l'information portée par plusieurs polymorphismes et de diminuer d'autre part le nombre de degrés de liberté du test. Arab-Alameddine et al. (2009) ont calculé un score d'activité basé sur le nombre d'allèles fonctionnels (combinaison de plusieurs *single nucleotide polymorphisms*) puis ils ont modélisé l'effet du polymorphisme sur la clairance apparente de l'efavirenz comme une fonction de la racine

carré de ce score.

Nous avons mis en évidence dans cette étude une association positive entre la clairance à la créatinine et la clairance apparente de la névirapine. Sur l'étendue des clairances à la créatinine observées, la variation de la clairance apparente de la névirapine est d'environ 27%. Cette association a aussi été mise en évidence dans l'étude de Gandhi et al. (2009) sur les 225 femmes de la *Women's Interagency HIV Study* aux États-Unis. La névirapine est essentiellement éliminée par voie métabolique mais Gandhi et al. (2009) expliquent cette association par l'effet des toxines urémiques sur les transporteurs et enzymes métaboliques du foie impliqués dans l'élimination de la névirapine.

L'hépatotoxicité précoce de la névirapine est encore un sujet débattu (Haas et al., 2006; Dailly et al., 2004). Dans leur étude, Gandhi et al. (2009) ont également mis en évidence une association positive entre des taux d'ALAT deux fois supérieurs à la normale et une exposition élevée à la névirapine chez des patientes sous traitement depuis au moins 6 mois. Dans l'essai PECAN, l'ensemble des patients étaient sous traitement depuis au moins 6 mois, certains étant passés de l'efavirenz à la névirapine après 12 mois de traitement. En conséquence, il serait intéressant de considérer la relation entre l'exposition à la névirapine et les marqueurs de l'activité hépatique tel que le taux d'ASAT, d'ALAT, de γGT et la bilirubinémie. Il est à noter que les co-traitements notamment la stavudine peuvent aussi entrainer une toxicité mitochondriale (McComsey et al., 2008).

Cette étude illustre dans quelle mesure les modèles non linéaires à effets mixtes permettent d'intégrer l'information recueillie chez l'ensemble des patients y compris lorsqu'un seul prélèvement est réalisé. Avec l'algorithme SAEM, l'approche par permutation a pu être utilisée sans difficulté dans ce contexte et a permis d'obtenir pour l'effet du génotype le moins représenté une p-value qui s'est avérée être deux fois plus grande que la valeur obtenue avec le test de Wald classique.

Chapitre 4

Conclusion générale

L'ensemble des méthodes évaluées dans ce manuscrit repose sur les MNLEM. Or l'utilisation de ces modèles si elle est en pleine expansion, reste encore minoritaire dans le domaine de la pharmacogénétique. Ils nécessitent par ailleurs d'avoir recours à des algorithmes d'estimation particulièrement sophistiqués (Pillai et al., 2005). Dans ces travaux nous avons utilisé trois de ces algorithmes, deux algorithmes basés sur une linéarisation de la vraisemblance et une méthode exacte. Au début de cette thèse, les algorithmes FO et FOCE étaient classiquement utilisés en pharmacocinétique de population alors que l'algorithme SAEM, en cours de développement, était encore très peu utilisé.

Notre étude présentée dans la partie 2.1 confirme les précédents résultats observés avec l'algorithme FO quant à la forte inflation de l'erreur de type I du test de Wald et du LRT (White et al., 1992; Gobburu et Lawrence, 2002; Wählby et al., 2001). Pourtant, l'algorithme FO est encore fréquement utilisé du fait des problèmes de convergence rencontrés avec l'algorithme FOCE. L'algorithme SAEM, dans notre étude de simulation présentée à la partie 2.2, obtient des résultats équivalents à l'algorithme FOCE en termes de biais, de précision et de performance des tests sur le plan d'expérience avec 40 sujets mais sans rencontrer de difficultés numériques. Théoriquement l'algorithme SAEM se comporte mieux qu'une méthode basée sur une linéarisation, mais sur le modèle pharmacocinétique que nous avons simulé, l'approximation réalisée par FOCE n'est pas mise en défaut. Cependant, l'instabilité numérique de l'algorithme FOCE est une limite importante à son utilisation et pour cette raison nous n'avons pas pu l'utiliser pour évaluer les tests de permutation. De plus, sur des modèles plus complexes comme ceux utilisés dans la partie 3.2 de ce manuscrit, l'algorithme FOCE ne permet pas forcément d'obtenir les erreurs d'estimation des paramètres du modèle.

Dans ce travail, nous avons évalué les performances de l'ANOVA, du test de Wald et du LRT pour détecter un effet gène sur un paramètre pharmacocinétique. Seule l'ANOVA conserve une erreur de type I non significativement différente de 5% quel que soit l'algorithme d'estimation. Étonnamment, les différents plans d'expérience considérés

dans nos travaux n'ont pas réussi à mettre cette méthode en défaut. D'autres études sur des protocoles avec des nombres de prélèvements plus variables et/ou des modèles plus complexes devraient permettre de mettre en évidence les limites de cette approche. En effet, l'ANOVA repose sur les estimations bayésiennes empiriques ce qui la rend sensible au phénomène de réduction vers la moyenne observé dans les protocoles non informatifs ou avec très peu de prélèvements par sujet. Pourtant en pharmacocinétique de population, l'effet de ce phénomène n'a été mis en évidence que dans les études en cross-over (Panhard et Mentré, 2005; Panhard et al., 2007; Dubois et al., 2009). Si l'ANOVA est essentiellement utilisée de nos jours pour le criblage des covariables, nous avons dans notre premier travail présenté dans la partie 2.1 défini une stratégie de sélection de modèles basée sur ce test. L'extension à x covariables génétiques est assez directe : après inclusion de la première covariable dans le modèle les effets aléatoires, et non plus les paramètres individuels, seraient comparés entre les différents génotypes des x-1 polymorphismes. Cependant, cette approche ne permet de considérer qu'un seul type de modèle de l'effet gène. Il n'est pas possible, par exemple, de considérer un modèle non linéaire où l'effet du gène est une fonction de la racine carré du nombre d'allèles fonctionnels (Arab-Alameddine et al., 2009). Un tel modèle permet pourtant de diminuer drastiquement le nombre de coefficients d'effet estimés.

Le test de Wald et le LRT obtiennent une erreur de type I significativement différente de 5% sur les trois plans d'expérience évalués dans l'étude de simulation présentée dans la partie 2.2. La plupart des études pharmacogénétiques réalisées aujourd'hui ne semblent donc pas pouvoir assurer le seuil nominal de ces tests asymptotiques. Pour des modèles comparables en termes de nombre de paramètres et de variabilités inter-sujets et résiduelle, Panhard et Mentré (2005), Panhard et al. (2007), Gobburu et Lawrence (2002) et Wählby et al. (2001) n'observent pas d'inflation de l'erreur de type I du LRT sur des plans d'expérience de moins de 160 observations avec une balance similaire entre le nombre de sujets et de prélèvements. Il semble que la catégorisation en 3 classes avec le déséquilibre des classes entraîné par les proportions d'Hardy-Weinberg et/ou le modèle multiplicatif nécessitent un plus grand nombre de sujets par rapport à une covariable binaire comme le sexe. D'autant plus que nous avons simulé le plus faible déséquilibre possible entre les génotypes d'un SNP biallélique, c'est à dire des variants alléliques en fréquence égale (50-50).

Dans notre recherche, nous n'avons pas considéré le test du score qui repose sur le rapport de la jacobienne et de la hessienne du modèle sans covariable. Ce test n'a pas été, à notre connaissance, utilisé dans le cadre de l'analyse de données pharmacogénétiques. De plus, il se situe par construction entre le LRT et le test de Wald qui repose sur la hessienne du modèle avec la covariable. Ses performances devraient donc se situer à mi-chemin de celles observées ici pour le LRT et le test de Wald.

En ce qui concerne les stratégies de construction de modèle basées sur les critères de

sélection, notre étude présentée dans la partie 2.1 de ce manuscrit rejoint globalement les conclusions des précédents travaux réalisés dans le cadre des modèles linéaires généralisés (GLM) (Burnham et Anderson, 2002). Néanmoins, le BIC dans notre étude se comporte correctement sans montrer de tendance conservatrice, comme il a pu être observé dans dans le cadre des GLM. Le résultat principal de cette évaluation est la contre-performance du critère d'Akaike et de son alternative développée pour les effectifs restreints, qui nous amène à ne pas recommander l'utilisation de ces deux critères dans le cadre des MNLEM. Cependant, d'autres études sont nécessaires pour pouvoir faire de réelles recommandations sur l'utilisation des autres critères évalués dans ce travail.

L'augmentation légère avec FOCE et SAEM mais significative de l'erreur de type I du test de Wald et du LRT nous a amené à considérer deux méthodes alternatives pour détecter un effet gène sur un paramètre pharmacocinétique, présentées dans la partie 2.3. Tout d'abord une approche non-paramétrique : le test de permutation, qui nécessite peu d'hypothèses et dont le principe est relativement intuitif (Manly, 1998). Ensuite nous nous sommes intéressés, assez logiquement après notre analyse de l'influence du design, aux distributions de Fisher approximées qui tiennent compte de l'éloignement à l'asymptotique dans le calcul du degré de liberté au dénominateur (Vonesh et Chinchilli, 1997). Le test de permutation s'avère une alternative convenable aux tests asymptotiques dans la mesure ou la méthode d'estimation ne rencontre pas de problèmes numériques pour des temps de calcul raisonnables, comme l'algorithme SAEM. Parmi les différentes méthodes basées sur une distribution F-approximée que nous avons évaluées, seule la méthode de Gallant (1975) corrige efficacement l'inflation de l'erreur de type I du test de Wald en pondérant la variance d'estimation en prenant en compte le nombre de sujet et le nombre d'effets fixes. Cette méthode n'a pas été développée dans le cadre des modèles mixtes au contraire de la méthode développée par Fai et Cornelius (1996) qui prend mieux en compte la nature longitudinale des données pharmacocinétiques dans le calcul du degré de liberté au dénominateur sans pour autant corriger l'inflation du test de Wald dans notre étude.

Nous avons appliqué les conclusions de nos différents travaux méthodologiques dans le cadre des traitements antirétroviraux VIH. Tout d'abord, nous avons analysé l'influence de la pharmacogénétique sur la pharmacocinétique de l'indinavir chez des patients naïfs de traitement par inhibiteur de protéases. Dans cette étude présentée dans la partie 3.1, un effet du variant *CYP* 3A4*1B est mis en évidence sur l'absorption de l'indinavir, possiblement à travers une diminution du pouvoir inhibant du ritonavir chez les homozygotes pour cet allèle. Ces données ayant inspiré l'étude de simulation réalisée pour l'évaluation des tests, nous avons utilisé l'approche par permutation pour calculer la p-value associée aux covariables conservées dans le modèle final. Dans un second temps, nous avons pu participer à l'analyse de concentrations de névirapine recueillies

en plusieurs occasions chez des patients cambodgiens décrite dans la partie 3.3. Comme les concentrations étaient recueillies à plusieurs occasions nous avons aussi pu calculer le coefficient de part génétique de la variabilité (R_{GC}) sur la clairance. Ce coefficient était d'environ 75%, c'est à dire qu'une part importante de la variabilité de la clairance de la névirapine pouvait être expliquée par des covariables d'origine génétique. Nous avons en effet montré dans cette étude l'influence du variant *CYP* 2B6*6 sur la pharmacocinétique de la névirapine. Ce variant est par ailleurs plus souvent rencontré dans la population cambodgienne que dans la population caucasienne. De nouveau, nous avons utilisé le test de permutation car le plan d'expérience correspondait à un des cas d'éloignement à l'asymptotique observé dans notre étude de simulation. Dans ces deux essais, les groupes de génotypes étaient déséquilibrés et le nombre de sujets dans l'essai COPHAR2 ainsi que le nombre de prélèvements par sujet dans l'essai PECAN étaient insuffisants pour assurer les conditions asymptotiques requises par le LRT et le test de Wald. Des corrections étaient donc nécessaires.

Dans le cadre d'une collaboration avec le département de pharmacocinétique des laboratoires de Recherche Servier, nous avons pu analyser les données d'une étude de phase II présentée dans la partie 3.2. Nous avons modélisé conjointement les concentrations du produit parent et de son métabolite actif par un modèle complexe, incluant un effet de premier passage et un phénomène de rétro-conversion, avec les algorithmes FOCE et SAEM. Cette fois-ci le R_{GC} a pu être calculé pour plusieurs paramètres du modèle et s'est montré particulièrement performant. En effet, nous avons mis en évidence l'influence du polymorphisme du *CYP* 2D6 sur la clairance d'élimination du métabolite, qui était le paramètre avec le R_{GC} le plus élevé. Comme les génotypes étaient très déséquilibrés nous avons calculé les p-values des covariables dans le modèle final par permutation. En ce qui concerne l'utilisation de MLXTRAN pour les modèles complexes, des analyses complémentaires sont nécessaires pour pouvoir déterminer l'origine des disparités observées dans les estimations des paramètres.

Une première extension de nos travaux serait de pouvoir déterminer lorsque le plan d'expérience satisfait ou non les conditions requises pour l'application des tests asymptotiques, ce qui permettrait d'éviter les tests de permutation. Cela pourrait être réalisé par simulations mais serait alors peu généralisable. Le principe soutenant une correction basée sur une distribution de Fisher pourrait aussi être exploré dans une optique plus centrée sur la quantification de la distance à l'asymptotique. Cette distance pourrait être quantifiée à travers l'écart entre les distributions théorique et corrigée par un calcul adéquat du degré de liberté au dénominateur ou encore à travers la pondération utilisée pour corriger la sous-évaluation de la matrice d'estimation. Dans ce cas, nous pourrions non seulement évaluer la distance à l'asymptotique mais peut-être aussi proposer une méthode pour optimiser les plans d'expérience afin d'obtenir de meilleures puissances

tout en assurant un seuil nominal de 5%.

Dans ses travaux sur les essais pharmacocinétiques en cross-over, Xavière Panhard a développé le test de Wald dans le cadre de l'équivalence (Panhard et Mentré, 2005; Panhard et al., 2007). Elle a montré que l'erreur de type I de ce test était similaire à celle du test de Wald classique sur plusieurs plans d'expérience. Dans notre recherche, nous n'avons considéré que des tests de comparaison. Il serait intéressant de vérifier que nous observons la même similarité entre les approches de comparaison et d'équivalence dans les études pharmacogénétiques.

En outre, nos travaux portent sur la détection d'un seul effet gène sur un unique paramètre pharmacocinétique. Deux autres perspectives, à plus long terme, seraient donc de considérer un effet gène sur plusieurs paramètres et/ou l'effet de plusieurs gènes. Sur ce deuxième point, plusieurs méthodes ont été développées en statistique génétique qui reposent sur le déséquilibre de liaison et proposent de capturer l'information de polymorphismes non génotypés. Une première approche, l'inférence haplotypique, considère les haplotypes comme l'unité de mesure de la variation génétique et est particulièrement puissante pour capturer les effets combinés de variants multiples fortement liés en *cis* (Clark, 2004). Cependant, il existe un certain nombre de problématiques posées par l'inférence haplotypique. Tout d'abord, les haplotypes ne sont pas directement observés et il faut donc inférer sur des données de génotypes non "phasées". Ensuite, les bornes de l'haplotype doivent être choisies et il est préférable qu'elles le soient dans des régions de forte recombinaison où le déséquilibre de liaison est nécessairement plus faible. Enfin, des stratégies de regroupement doivent être considérées pour gérer le nombre d'haplotypes dont certains seront rares qui contribueront au nombre de degrés de liberté du test et diminueront sa puissance. Une seconde approche, l'inférence multipoint, utilise les SNPs comme unité de mesure et augmente le nombre de SNPs observés par imputation (Marchini et al., 2007). Ces imputations sont réalisées en s'appuyant sur les grandes cohortes de génotypages comme celles construites dans le cadre du projet Hapmap (http://www.hapmap.org/). Il serait intéressant d'évaluer ces différentes méthodes dans les MNLEM en vue du passage de la stratégie gène candidat à des analyses plus vastes voire génome entier en pharmacocinétique de population.

La pharmacogénétique représente une voie d'accès importante vers l'individualisation des traitements. Une meilleure efficacité et la maîtrise des effets indésirables ne sont pas des enjeux mineurs, notamment dans des pathologies chroniques comme le VIH ou le VHB. Mais peu d'études qui explorent les polymorphismes associés à la réponse pharmacocinétique et/ou pharmacodynamique suivent les recommandations des autorités de santé. Ces dernières préconisent d'augmenter le nombre de patients inclus et nous avons montré que cela ne devait pas être fait au détriment de l'information recueillie chez chaque sujet. Dans cette thèse, nous espérons avoir démontré que les MNLEM sont un outil particulièrement adapté à la planification et l'évaluation de telles études qui

permettront l'avancée de la recherche dans ce domaine de santé publique.

Bibliographie

1. Akaike H., A new look at the statistical model identification. *IEEE Trans Automat Contr*, 1974, vol. 19, p. 716–723

2. Arab-Alameddine M., Di Iulio J., Buclin T., Rotger M., Lubomirov R., Cavassini M., Fayet A., Décosterd L.A., Eap C.B., Biollaz J., Telenti A., Csajka C., the Swiss HIV Cohort Study, Pharmacogenetics-based population pharmacokinetic analysis of efavirenz in HIV-1-infected individuals. *Clin Pharmacol Ther*, 2009, vol. 85, p. 485–494

3. Balram C., Zhou Q., Cheung Y.B., Lee E.J., CYP3A5*3 and *6 single nucleotide polymorphisms in three distinct asian populations. *Eur J Clin Pharmacol*, 2003, vol. 59, p. 123–126

4. Barton N., Pleiotropic models of quantitative trait variation. *Genetics*, 1990, vol. 124, p. 773–782

5. Bertrand J., Mentré F., *Mathematical Expressions of the Pharmacokinetic and Pharmacodynamic Models implemented in the Monolix software*. MONOLIX group, Paris, France, 2008, http://software.monolix.org/download31b/Monolix31_PKPD_library.pdf

6. Bertrand J., Tréluyer J., Panhard X., Tran A., S, Rey E., Salmon-Céron D., Duval X., Mentré F., study group t., Influence of pharmacogenetics on indinavir disposition and short-term response in HIV patients initiating HAART. *Eur J Clin Pharmacol*, 2009, vol. 65, p. 667–678

7. Blanco J.G., Edick M.J., Hancock M.L., Winick N.J., T D., Amylon M.D., Bash R.O., Behm F.G., Camitta B.M., Pui C.H., Raimondi S.C., Relling M.V., Genetic polymorphisms in CYP3A5, CYP3A4 and NQO1 in children who developed therapy-related myeloid malignancies. *Pharmacogenetics*, 2002, vol. 12, p. 605–611

8. Bonate P., Covariate detection in population pharmacokinetics using partially linear mixed effects models. *Pharm Res*, 2005, vol. 22, p. 541–549

9. Bozdogan H., Model selection and Akaike's information criteria (AIC) : The general theory and its analytical extensions. *Psychometrika*, 1987, vol. 52, p. 345–370

10. Brendel K., Legrand M., Taburet A.M., Baron G., Goujard C., Mentré F., Population pharmacokinetic analysis of indinavir in HIV-infected patient treated with a stable antiretroviral therapy. *Fundam Clin Pharmacol*, 2005, vol. 19, p. 373–383

11. Burnham K.P., Anderson D.R., *Model Selection and Multimodel A Practical Information - Theoric Approach*. New York : Springer-Verlag, 2002

12. Cheeseman S.H., Havlir D., McLaughlin M.M., Greenough T.C., Sullivan J.L., Hall D., Hattox S.E., Spector S.A., Stein D.S., Myers M., Phase i/ii evaluation of nevirapine alone and in combination with zidovudine for infection with human immunodeficiency virus. *J Acquir Immune Defic Syndr Hum Retrovirol*, 1995, vol. 8, p. 141–151

13. Chen B.L., Chen Y., Tu J.H., Li Y.L., Zhang W., Li Q., Fan L., Tan Z.R., Hu D.L., Wang D., Wang L.S., Ouyang D.S., Zhou H.H., Clopidogrel inhibits CYP2C19-dependent hydroxylation of omeprazole related to CYP2C19 genetic polymorphisms. *J Clin Pharmacol*, 2009, vol. 49, p. 574–581

14. Clark A.G., The role of haplotypes in candidate gene studies. *Genet Epidemiol*, 2004, vol. 27, p. 321–333

15. Collins F.S., Morgan M., Patrinos A., The human genome project : Lessons from large-scale biology. *Science*, 2003, vol. 300, p. 286–290

16. Comets E., Mentré F., Evaluation of tests based on individual versus population modeling to compare dissolution curves. *J Biopharm Stat*, 2001, vol. 11, p. 107–123

17. Comets E., Verstuyft C., Lavielle M., Jaillon P., Becquemont L., Mentré F., Modelling the influence of MDR1 polymorphism on digoxin pharmacokinetic parameters. *Eur J Clin Pharmacol*, 2007, vol. 63, p. 437–449

18. Commenges D., G H.J., Proust C., Guedj J., A newton-like algorithm for likelihood maximization : The robust-variance scoring algorithm. *arXiv :math/0610402v2*, 2005

19. Committee for human medicinal products, Reflection paper on the use of pharmacogenetics in the pharmacokinetic evaluation of medicinal products. Technical report, EMEA, 2007, http://www.emea.europa.eu/pdfs/human/pharmacogenetics/20191406en.pdf

20. Crow J.F., Hardy, Weinberg and language impediments. *Genetics*, 1999, vol. 152, p. 821–825

21. Csajka C., Marzolini C., Fattinger K., Decosterd L.A., Telenti A., Biollaz J., Buclin T., Population pharmacokinetics of indinavir in patients infected with

human immunodeficiency virus. *Antimicrob Agents Chemother*, 2004, vol. 48, p. 3226–3232

22. Dailly E., Billaud E., Reliquet V., Breurec S., Perré P., Léautez S., Jolliet P., Bourin M., Raffi F., No relationship between high nevirapine plasma concentration and hepatotoxicity in HIV-1-infected patients naive of antiretroviral treatment or switched from protease inhibitors. *Eur J Clin Pharmacol*, 2004, vol. 60, p. 343–348

23. Daly A.K., Steen V.M., Fairbrother K.S., Idle J.R., CYP2D6 multiallelism. *Methods Enzymol*, 1996, vol. 272, p. 199–210

24. Daly M.J., Rioux J.D., Schaffner S.F., Hudson T.J., Lander E.S., High-resolution haplotype structure in the human genome. *Nat Genet*, 2001, vol. 29, p. 229–232

25. Dubois A., Gsteiger S., Pigeolet E., Mentré F., Bioequivalence tests based on individual estimates using non-compartmental or model-based analyses : evaluation of estimates of sample means and type I error for different designs. *Pharm Res*, 2009

26. Duval X., Mentré F., Rey E., Auleley S., Peytavin G., Biour M., Métro A., Goujard C., Taburet A.M., Lascoux C., Panhard X., Tréluyer J., Salmon D., Benefit of therapeutic drug monitoring of protease inhibitors in HIV-infected patients depends on PI used in HAART regimen-ANRS 111 trial. *Fundam Clin Pharmacol*, 2009, vol. 23, p. 491–500

27. Erickson D.A., Mather G., Trager W.F., Levy R.H., Keirns J.J., Characterization of the in vitro biotransformation of the HIV-1 reverse transcriptase inhibitor nevirapine by human hepatic cytochromes P-450. *Drug Metab Dispos*, 1999, vol. 27, p. 1488–1495

28. Evans D.A., Harmer D., Downham D.Y., Whibley E.J., Idle J.R., Ritchie J., Smith R.L., The genetic control of sparteine and debrisoquine metabolism in man with new methods of analysing bimodal distributions. *J Med Genet*, 1983, vol. 20, p. 321–329

29. Fai A.H.T., Cornelius P.L.C., Approximate f-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *J Stat Comput Simul*, 1996, vol. 54, p. 363–378

30. FDA, ICH topic E15 definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories. Technical report, FDA, 2008, http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm129296.pdf

31. Fellay J., Marzolini C., Meaden E., Back D., Buclin T., Chave J., Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter gene MDR1 : a pharmacogenetic study. *Lancet*, 2002, vol. 359, p. 30–36

32. Ferradini L., Laureillard D., Prak N., Ngeth C., Fernandez M., Pinoges L., Puertas G., Taburet A.M., Ly N., Rouzioux C., Balkan S., Quillet C., Delfraissy J.F., Positive outcomes of HAART at 24 months in HIV-infected patients in cambodia. *AIDS*, 2007, vol. 21, p. 2293–2301

33. for human medicinal products C., ICH topic E15 definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories. Technical report, EMEA, 2008, http://www.emea.europa.eu/pdfs/human/ich/43798606en.pdf

34. Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A., Faggart M., Liu-Cordero S.N., Rotimi C., Adeyemo A., Cooper R., Ward R., Lander E.S., Daly M J and A.D., The structure of haplotype blocks in the human genome. *Science*, 2002, vol. 296, p. 2225–2229

35. Gabrielsson J., Weiner D., *Pharmacokinetic and Pharmacodynamic Data Analysis : Concepts and Applications*. Stockholm : Apotekarsocieteten, 1999

36. Gallant A.R., Seemingly unrelated nonlinear regressions. *J Econom*, 1975, vol. 3, p. 35–50

37. Gandhi M., Benet L.Z., Bacchetti P., Kalinowski A., Anastos K., Wolfe A.R., Young M., Cohen M., Minkoff H., Gange S.J., Greenblatt R.M.W., Nonnucleoside reverse transcriptase inhibitor pharmacokinetics in a large unselected cohort of HIV-infected women. *J Acquir Immune Defic Syndr*, 2009, vol. 50, p. 482–491

38. Ge Z., Bickel P., Rice J., An approximate likelihood approach to nonlinear mixed effects models via spline approximation. *Comput Stat Data Anal*, 2004, vol. 46, p. 747–776

39. Gibaldi M., Perrier D., *Pharmacokinetics*. Marcel Dekker, New York, 2 ed., 1982

40. Gobburu J., Lawrence J., Application of resampling techniques to estimate exact significance levels for covariate selection during nonlinear mixed effects model building : some inferences. *Pharm Res*, 2002, vol. 19, p. 92–98

41. Haas D.W., Bartlett J.A., Andersen J.W., Sanne I., Wilkinson G.R., Hinkle J., Rousseau F., Ingram C.D., Shaw A., Lederman M.M., Kim R.B.A.,

Pharmacogenetics of nevirapine-associated hepatotoxicity : an adult AIDS clinical trials group collaboration. *Clin Infect Dis*, 2006, vol. 43, p. 783–786

42. Haas D.W., Gebretsadik T., Mayo G., Menon U.N., Acosta E.P., Shintani A., Floyd M., Stein C.M., Wilkinson G.R., Associations between CYP2B6 polymorphisms and pharmacokinetics after a single dose of nevirapine or efavirenz in african americans. *J Infect Dis*, 2009, vol. 199, p. 872–880

43. Haas D.W., Ribaudo H.J., Kim R.B., Tierney C., Wilkinson G.R., Gulick R.M., Clifford D.B., Hulgan T., Marzolini C., Acosta E.P., Pharmacogenetics of efavirenz and central nervous system side effects : an adult AIDS clinical trials group study. *AIDS*, 2004, vol. 18, p. 2391–2400

44. Han J.Y., Lim H.S., Park Y.H., Lee S.Y., Lee J.S., Integrated pharmacogenetic prediction of irinotecan pharmacokinetics and toxicity in patients with advanced non-small cell lung cancer. *Lung Cancer*, 2009, vol. 63, p. 115–120

45. Hinds D.A., Stuve L.L., Nilsen G.B., Halperin E., Eskin E., Ballinger D.G., Frazer K.A., Cox D.R., Whole-genome patterns of common DNA variation in three human populations. *Science*, 2005, vol. 307, p. 1072–1079

46. Hirt D., Mentré F., Tran A., Rey E., Auleley S., Salmon D., Duval X., Tréluyer J.M., the COPHAR 2-ANRS Study Group, Effect of CYP2C19 polymorphism on nelfinavir to M8 biotransformation in HIV patients. *Br J Clin Pharmacol*, 2008, vol. 65, p. 548–57

47. International HapMap Consortium, The international hapmap project. *Nature*, 2003, vol. 426, p. 789–796

48. Kalow W., Meyer U.A., Tyndale R.F., eds., *Pharmacogenomics*. New York : Marcel Dekker, 2001

49. Kappelhoff B.S., Huitema A.D.R., Sankatsing S.U.C., Meenhorst P.L., Van Gorp E.C.M., Mulder J.W., Prins J.M., Beijnen J.H., Population pharmacokinetics of indinavir alone and in combination with ritonavir in HIV-1-infected patients. *Br J Clin Pharmacol*, 2005a, vol. 60, p. 276–286

50. Kappelhoff B.S., R H.A.D., Crommentuyn K.M.L., Mulder J.W., Meenhorst P.L., van Gorp E.C.M., Mairuhu A.T.A., Beijnen J.H., Development and validation of a population pharmacokinetic model for ritonavir used as a booster or as an antiviral agent in HIV-1-infected patients. *Br J Clin Pharmacol*, 2005b, vol. 59, p. 174–182

51. Karlsson M.O., Holford N., A tutorial on visual predictive checks. *17th Population Approach Group in Europe*, Marseille, France, 2008 www.page-meeting.org/?abstract=1434

52. Kuhlmann J., Alternative strategies in drug development : clinical pharmacological aspects. *Int J Clin Pharmacol Ther*, 1999, vol. 37, p. 575–583

53. Kuhn E., Lavielle M., Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal*, 2005, vol. 49, p. 1020–38

54. Kullback S., Leibler R.A., On information and sufficiency. *Ann Math Stat*, 1951, vol. 22, p. 79–86

55. Lamba J.K., Lin Y.S., Schuetz E.G., Thummel K.E., Genetic contribution to variable human CYP3A-mediated metabolism. *Adv Drug Deliver Rev*, 2002, vol. 54, p. 1271–94

56. Lavielle M., *MONOLIX (MOdèles NOn LInéaires à effets miXtes)*. MONOLIX group, Orsay, France, 2008, http://software.monolix.org/sdoms/software/

57. Lee P., Design and power of a population pharmacokinetic study. *Pharm Res*, 2001, vol. 18, p. 75–82

58. Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L., Walenz B.P., Axelrod N., Huang J., Kirkness E.F., Denisov G., Lin Y., MacDonald J.R., Pang A.W., Shago M., Stockwell T.B., Tsiamouri A., Bafna V., Bansal V., Kravitz S.A., Busam D.A., Beeson K.Y., McIntosh T.C., Remington K.A., Abril J.F., Gill J., Borman J., Rogers Y.H., Frazier M.E., Scherer S.W., Strausberg R.L., Venter J.C., The diploid genome sequence of an individual human. *PLoS Biol*, 2007, vol. 5, p. e254

59. Licinio J., Wong M., eds., *Pharmacogenomics. The Search for Individualized Therapies*. Weinheim : WILEY-VCH, 2002

60. Lindstrom M., Bates D., Nonlinear mixed effects models for repeated measures data. *Biometrics*, 1990, vol. 46, p. 673–687

61. Long J.C., Kittles R.A., Human genetic diversity and the nonexistence of biological races. *Hum Biol*, 2003, vol. 75, p. 449–471

62. Manly B.F.J., *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Texts in statistical science, Chapman & Hall, London, 2 ed., 1998

63. Marchini J., Howie B., Myers S., McVean G., Donnelly P., A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*, 2007, vol. 39, p. 906–913

64. Marzolini C., Paus E., Buclin T., Kim R.B., Polymorphisms in human MDR1 (p-glycoprotein) : recent advances and clinical relevance. *Clin Pharmacol Ther*, 2003, vol. 75, p. 13–33

65. McComsey G.A., Lo Re V.r., O'Riordan M., Walker U.A., Lebrecht D., Baron E., Mounzer K., Frank I., Effect of reducing the dose of stavudine on body composition, bone density, and markers of mitochondrial toxicity in HIV-infected subjects : a randomized, controlled study. *Clin Infect Dis*, 2008, vol. 46, p. 1290–126

66. Moltó J., Valle M., Miranda C., Cedeño S., Miranda J., Santos J., Negredo E., Vilaró J., Costa J., Clotet B., Once- or twice-daily dosing of nevirapine in HIV-infected adults : a population pharmacokinetics approach. *J Antimicrob Chemother*, 2008, vol. 62, p. 784–792

67. Motulsky A., Drug reactions, enzymes and biochemical genetics. *JAMA*, 1957, vol. 165, p. 835–837

68. Nelson D.R., Koymans L., Kamataki T., Stegeman J.J., Feyereisen R., Waxman D.J., Waterman M.R., Gotoh O., Coon M.J., Estabrook R.W., Gunsalus I.C., Nebert D.W., P450 superfamily : update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, 1996, vol. 6, p. 1–42

69. Ng P.C., Zhao Q., Levy S., Strausberg R.L., Venter J.C., Individual genomes instead of race for personalized medicine. *Nature*, 2008, vol. 84, p. 306–309

70. Owen R.P., Sangkuhla K., Kleina T.E., Altmana R.B., Cytochrome P450 2D6. *Pharmacogenet Genomics*, 2009, vol. 19, p. 559–562

71. Ozdemir V., Kalow W., Tang B., Paterson A.D., Walker S.E., Endrenyi L., Kashuba A.D.M., Evaluation of the genetic component of variability in CYP3A4 activity : a repeated drug administration method. *Pharmacogenetics*, 2000, vol. 10, p. 373–88

72. Panhard X., Mentré F., Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence cross-over trials. *Stat Med*, 2005, vol. 24, p. 1509–1524

73. Panhard X., Samson A., Extension of the SAEM algorithm for nonlinear mixed models with 2 levels of random effects. *Biostatistics*, 2009, vol. 10, p. 121–135

74. Panhard X., Taburet A., Piketti C., Mentré F., Impact of modelling intra-subject variability on tests based on non-linear mixed-effects models in cross-over pharmacokinetic trials with application to the interaction of tenofovir on atazanavir in HIV patients. *Stat Med*, 2007, vol. 26, p. 1268–1284

75. Paskind H.A., Some differences in response to atropine in white and coloured races. *J Am Med Assoc*, 1921, vol. 76, p. 104–108

76. Patin E., Barreiro L.B., Sabeti P.C., Austerlitz F., Luca F., Sajantila A., Behar D.M., Semino O., Sakuntabhai A., Guiso N., Gicquel B., McElreavey K., Harding R.M., Heyer E., Quintana-Murci L., Deciphering the ancient and complex evolutionary history of human arylamine n-acetyltransferase genes. *Am J Hum Genet*, 2006, vol. 78, p. 423–436

77. Penzak S.R., Kabuye G., Mugyenyi P., Mbamanya F., Natarajan V., Alfaro R.M., Kityo C., Formentini E., Masur H., Cytochrome P450 2B6 (CYP2B6) G516T influences nevirapine plasma concentrations in HIV-infected patients in Uganda. *HIV Med*, 2007, vol. 8, p. 86–91

78. Pillai G.C., Mentré F., Steimer J.L., Non-linear mixed effects modeling - from methodology and software development to driving implementation in drug development science. *J Pharmacokinet Pharmacodyn*, 2005, vol. 32, p. 161–183

79. Pinheiro J.C., Bates D.M., *Mixed Effects Models in S and S-Plus*. Springer Verlag, New York, 2000

80. Quaranta S., Chevalier D., Allorge D., Lo-Guidice J.M., Migot-Nabias F., Kenani A., Imbenotte M., Broly F., Lacarelle B., Lhermitte M., Ethnic differences in the distribution of CYP3A5 gene polymorphisms. *Xenobiotica*, 2006, vol. 36, p. 1191–1200

81. R Development Core Team, *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008

82. Raftery A.E., Bayesian model selection in social research (with discussion). *Sociol Methodol*, 1995, vol. 25, p. 111–196

83. Retout S., Comets E., Nagard H.L., Bazzoli C., Mentré F., *PFIM Interface 21*. UMR738, INSERM, Université Paris 7, Paris, France, 2007a, http://www.pfim.biostat.fr/

84. Retout S., Comets E., Samson A., Mentré F., Design in nonlinear mixed effects models : optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates. *Stat Med*, 2007b, vol. 26, p. 5162–5179

85. Riska P., Lamson M., MacGregor T., Sabo J., Hattox S., Pav J., Keirns J., Disposition and biotransformation of the antiretroviral drug nevirapine in humans. *Drug Metab Dispos*, 1999, vol. 27, p. 895–901

86. Rotger M., Colombo S., Furrer H., Bleiber G., Buclin T., Lee B.L., Keiser O., Biollaz J., Décosterd L., Telenti A.S., Influence of CYP2B6 polymorphism on plasma and intracellular concentrations and toxicity of efavirenz and nevirapine in HIV-infected patients. *Pharmacogenet Genomics*, 2005, vol. 15, p. 1–5

87. Sakaeda T., Nakamura T., Okumura K., MDR1 genotype-related pharmacokinetics and pharmacodynamics. *Biol Pharm Bull*, 2002, vol. 25, p. 1391–1400

88. SAS Institute Inc, *SAS 912*. Cary, NC, USA, 2004

89. Schwartz G., Estimating the dimension of a model. *Ann Stat*, 1978, vol. 6, p. 461–464

90. Sheiner L., Beal S., *NONMEM Version 51*. University of California, NONMEM Project Group, San Francisco, 1998

91. Sheiner L., Rosenberg B., Melmon K.L., Modelling of individual pharmacokinetics for computer aided drug dosage. *Computers and Biomedical Research*, 1972, vol. 5, p. 441–459

92. Solas C., Simon N., Drogoul N., Quaranta S., Frixon-Marin V., Brunet V.B.R.C., Gastaut J.A., Durand A., Lacarelle B., Poizot-Martin I., Minimal effect of MDR1 and CYP3A5 genetic polymorphisms on the pharmacokinetics of indinavir in HIV-infected patients. *Br J Clin Pharmacol*, 2007, vol. 64, p. 353–362

93. Spire B., Carrieri P., Sopha P., Protopopescu C., Prak N., Quillet C., Ngeth C., Ferradini L., Delfraissy J.F., Laureillard D., Adherence to antiretroviral therapy in patients enrolled in a comprehensive care program in Cambodia : a 24-month follow-up assessment. *Antivir Ther*, 2008, vol. 13, p. 697–703

94. Strachan T., Read A.P., *Human Molecular Genetics 2*. Oxford : Wiley, 1999

95. Sugiura N., Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun Stat Theory Methods*, 1978, vol. 7, p. 13–26

96. Tanigawara Y., Aoyama N., Kita T., Shirakawa K., Komada F., Kasuga M., Okumura K., CYP2C19 genotype-related efficacy of omeprazole for the treatment of infection caused by helicobacter pylori. *Clin Pharmacol Ther*, 1999, vol. 66, p. 528–534

97. Tong K., He M.L., Lin C.K., Guo L., Kung H.F., Sung J.J., Lee S.S., The implications of a high allelic frequency of CYP2B6 g516t in ethnic chinese persons. *Clin Infect Dis*, 2006, vol. 43, p. 541–542

98. Vonesh E.F., A note on the use of laplace's approximation for nonlinear mixed-effects models. *Biometrika*, 1996, vol. 83, p. 447–452

99. Vonesh E.F., Chinchilli V.M., *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York, 1997

100. Werner D., Werner U., Meybaum A., Schmidt B., Umbreen S., Grosch A., Lestin H.G., Graf B., Zolk O., Fromm M.F., Integrated pharmacogenetic prediction of irinotecan pharmacokinetics and toxicity in patients with advanced non-small cell lung cancer. *Clin Pharmacokinet*, 2008, vol. 47, p. 323–332

101. White D., Walawander C., Liu D., Grasela T., Evaluation of hypothesis testing for comparing two populations using NONMEM analysis. *J Pharmacokinet Biopharm*, 1992, vol. 20, p. 295–313

102. Wählby U., Jonsson E., Karlsson M., Assessment of actual significance levels for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn*, 2001, vol. 28, p. 231–252

103. Wilson J.F., Weale M.E., Smith A.C., Gratrix F., Fletcher B., Thomas M.G., Bradman N., Goldstein D.B., Population genetic structure of variable drug response. *Nat Genet*, 2001, vol. 29, p. 265–269

104. Wolfinger R., Laplace's approximation for nonlinear mixed models. *Biometrika*, 1993, vol. 80, p. 791–795

105. Yamasaki Y., Ieiri I., Kusuhara H., Sasaki T., Kimura M., Tabuchi H., Ando Y., Irie S., Ware J.A., Nakai Y., Higuchi S., Sugiyama Y., Pharmacogenetic characterization of sulfasalazine disposition based on NAT2 and ABCG2 (BCRP) gene polymorphisms in humans. *Clin Pharmacol Ther*, 2008, vol. 84, p. 95–103

106. Yang J.C., Yang Y.F., Uang Y.S., Lin C.J., Wang T.H., Pharmacokinetic-pharmacodynamic analysis of the role of CYP2C19 genotypes in short-term rabeprazole-based triple therapy against helicobacter pylori. *Br J Clin Pharmacol*, 2009, vol. 67, p. 503–510

107. Yeni P., groupe des experts "Prise en charge médicale des personnes infectées par le VIH", Prise en charge médicale des personnes infectées par le vih. Technical report, 2006, http://www.sante-sports.gouv.fr/IMG//pdf/rapport_experts_2006.pdf