



HAL
open science

Modélisation de données longitudinales : protocoles robustes dans les essais cliniques, interactions bactériophages/bactéries

Jérémy Seurat

► **To cite this version:**

Jérémy Seurat. Modélisation de données longitudinales : protocoles robustes dans les essais cliniques, interactions bactériophages/bactéries. Statistiques [math.ST]. Université Paris Cité, 2021. Français. NNT : 2021UNIP5230 . tel-04551744

HAL Id: tel-04551744

<https://theses.hal.science/tel-04551744>

Submitted on 18 Apr 2024

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

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

Université de Paris

École doctorale 393 Pierre Louis de santé publique

Épidémiologie et sciences de l'information biomédicale

*UMR 1137 – IAME – Infection, Antimicrobiens, Modélisation, Évolution,
INSERM, Université de Paris*

*Équipe BIPID – Modélisation biostatistique, pharmacométrie, et investigation
clinique en maladies infectieuses*

Modélisation de données longitudinales : protocoles robustes dans les essais cliniques, interactions bactériophages/bactéries

Par Jérémy Seurat

Thèse de doctorat de biostatistiques et biomathématiques

Dirigée par le Pr. France Mentré, co-encadrée par le Dr. Thu Thuy Nguyen

Présentée et soutenue publiquement le 8 mars 2021

Devant un jury composé de :

Florence Gattacceca,	MCU, Université d'Aix-Marseille,	examineur
Nicolas Grégoire,	MCU, Université de Poitiers,	rapporteur
Vincent Jullien,	PU-PH, Université Sorbonne Paris Nord,	examineur
Estelle Kuhn,	DR, Université Paris-Saclay,	rapporteur
France Mentré,	PU-PH, Université de Paris,	directeur de thèse
Thu Thuy Nguyen,	Dr., Novartis,	co-encadrant
Luc Pronzato,	DR, Université Côte d'Azur,	examineur



Except where otherwise noted, this work is licensed under
<https://creativecommons.org/licenses/by-nd/3.0/fr/>

Remerciements

Cette thèse fut remplie de moments scientifiquement et humainement enrichissants. Je remercie les personnes rencontrées sur le chemin menant au diplôme.

A Estelle Kuhn et Nicolas Grégoire qui m'ont fait l'honneur d'être rapporteurs de ma thèse, pour leur lecture attentive du manuscrit et leurs suggestions.

A Florence Gattacceca, Vincent Jullien et Luc Pronzato qui m'ont fait l'honneur d'examiner ce travail et de faire partie de mon jury de thèse.

A France, qui a accepté de diriger cette thèse. Ta rigueur, ta volonté de transmettre, de faire progresser dans les différents aspects scientifiques et du monde de la recherche sont une source d'inspiration. Je te remercie aussi d'avoir pu être si disponible, y compris ces derniers mois, d'avoir fait preuve de pragmatisme et d'humanité tout au long de ces années de doctorat.

A Thu Thuy, qui a accepté de m'encadrer en stage de Master 2 puis de co-encadrer cette thèse. Je te remercie pour ton extrême gentillesse, ton adaptabilité, tes connaissances et je te remercie surtout pour ta faculté à les expliquer. Cela a été essentiel pour bâtir les fondations de ma thèse.

A Jérémie, qui m'a encadré sur le travail de modélisation des données de phages, merci pour ta disponibilité, le partage de tes connaissances et ton esprit critique.

Aux personnes ayant participé et aidé aux différents travaux menés durant ma thèse, en particulier Yuxin, Hervé, Ivana, Romain, Pascal Girard, Kosalaram Goteti, Raphaëlle et Laurent.

Aux gestionnaires de IAME, merci beaucoup Houda de me faciliter la vie depuis mon arrivée dans l'unité ainsi que pour ton dynamisme au labo et en dehors, merci aussi à Tiphanie, Myriam et Vincent.

A l'équipe informatique/Catibiomed, sans qui aucun run Monolix ne serait possible, merci Hervé (encore !), Lionel, Karim, Saadia et François.

Aux étudiants ou chercheurs de IAME, passés ou présents, en particulier à François, Charles, Julie, Jimmy, Emmanuelle, Cédric, Niccolo, Tram, Florence, Emilie, Morgane (les deux), Pauline, Marc, Antonio, Vincent, Simon, Nathan, Claire, Marilou, Coralie, Selma, Mathieu, Manel, Jinju, Guillaume, Aurélien, Marion, Mélanie, Alexandra, Drifa, Arthur, Anne, Emmanuelle, Dara, Alice, Lucas, Nadia, Sylvie, Anthony, Nadège, Gaspard et Donia. Je vous remercie chaleureusement pour l'émulation et pour nos moments de franche camaraderie.

A mes amis (les vrais, ceux qui se reconnaitrons) d'enfance, du lycée, de pharma ou d'ailleurs, merci pour votre soutien et pour tous les moments partagés et à venir.

A ma famille, en particulier à mes chers parents. Merci pour votre éducation, votre amour, votre générosité et votre soutien sans faille.

A Laure-Hélène, merci d'enseigner ma vie.

Production scientifique dans le cadre de la thèse

Articles

Publications acceptées

Seurat J, Nguyen TT, Mentré F. Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes. *Stat Methods Med Res.* 2020 Mar;29(3):934–52.

Seurat J, Girard P, Goteti K, Mentré F. Comparison of various Phase I combination therapy designs in oncology for evaluation of early tumor shrinkage using simulations. *CPT Pharmacometrics Syst Pharmacol.* 2020 Dec;9(12):686-694.

Seurat J*, Tang Y*, Mentré F, Nguyen TT. Finding optimal design in nonlinear mixed effect models using multiplicative algorithms. *Comput Methods Programs Biomed.* In Press.

Publication soumise

Delattre R*, **Seurat J***, Haddad F, Nguyen TT, Gaborieau B, Kane R, Ricard JD, Debarbieux L, Guedj J. Deployment of phage therapy: from *in vivo* to *in silico* models for the treatment of bacterial pneumonia.

* : co-premier auteur

Communications orales

Loingeville F*, **Seurat J***, Nguyen TT, Rivière MK, Mentré F. Using Hamiltonian Monte-Carlo to design longitudinal studies with discrete outcomes accounting for parameter and model uncertainties. *Population Optimum Design of Experiments (PODE)*. Bâle, Suisse. 2017.

* : co-présentateur

Seurat J, Nguyen TT, Mentré F. Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes. *GDR « Statistiques et Santé »*. Paris, France. 2019.

Présentations affichées

Seurat J, Nguyen TT, Mentré F. Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes. *Design of experiments (DoE)*. Marseille, France. 2018.

Seurat J, Nguyen TT, Mentré F. Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes. *Population Approach Group Europe (PAGE)*. Montreux, Suisse. 2018.

Seurat J, Girard P, Sharma VD, Goteti K, Mentré F. Comparison by clinical trial simulation of various Phase I combination therapy designs in oncology for evaluation of early tumor shrinkage. *Population Approach Group Europe (PAGE)*. Stockholm, Suède. 2019.

Seurat J, Tang Y, Nguyen TT, Mentré F, Le Nagard H, on behalf of the PFIM group. New version of PFIM for optimal design in nonlinear mixed effects models using R S4. *Population Approach Group Europe (PAGE)*. Stockholm, Suède. 2019.

Liste des principales abréviations

- AGQ : *Adaptive Gaussian quadrature* (Quadrature de Gauss adaptative)
- AIC : *Akaike information criteria* (Critère d'information d'Akaike)
- AUC : *Area under the curve* (Aire sous la courbe)
- CFU : *Colony-forming unit* (Unité formant colonie)
- CTS : *Clinical trial simulation* (Simulation d'essais cliniques)
- EBE : *Empirical Bayes estimates* (Estimation empirique bayésienne)
- ETS : *Early tumor shrinkage* (Décroissance tumorale précoce)
- FIM : *Fisher information Matrix* (Matrice d'information de Fisher)
- FO : *First order* (Premier ordre)
- FW : Fedorov-Wynn
- HMC : *Hamiltonian Monte-Carlo* (Monte-Carlo Hamiltonien)
- MC : Monte-Carlo
- MNLEM : Modèle non-linéaire à effets mixtes
- NPDE : *Normalized predictive distribution error* (Résidus simulés normalisés décorrélés)
- PD : Pharmacodynamie
- PFU : *Plaque-forming unit* (Unité formant plaque)
- PK : *Pharmacokinetics* (Pharmacocinétique)
- QRMC : *Quasi random Monte-Carlo* (Monte-Carlo quasiment aléatoire)
- RSE : *Relative standard error* (Erreur standard relative)
- SAEM : *Stochastic Approximation Expectation Maximization*
- VPC : *Visual predictive check* (Vérification visuelle des prédictions)

Sommaire

PREMIÈRE PARTIE : INTRODUCTION 10

1.1.	MODÈLES NON-LINÉAIRES À EFFETS MIXTES	10
1.1.1.	MODÈLES POUR DONNÉES CONTINUES.....	13
1.1.2.	MODÈLES POUR DONNÉES NON CONTINUES.....	14
1.1.3.	MÉTHODES ET LOGICIELS D'ESTIMATION	16
1.1.4.	SÉLECTION ET ÉVALUATION DE MODÈLES.....	17
1.1.5.	MODÉLISATION ET ONCOLOGIE.....	19
1.2.	EVALUATION ET OPTIMISATION DE PROTOCOLES DANS LES MODÈLES MIXTES	20
1.2.1.	MATRICE D'INFORMATION DE FISHER	21
1.2.1.1.	Méthodes d'évaluations de la FIM.....	23
1.2.1.2.	Critères d'optimalité	24
1.2.1.3.	Incertitude du modèle et des paramètres	25
1.2.1.4.	Logiciels et algorithmes d'optimisation	26
1.2.2.	SIMULATIONS D'ESSAIS CLINIQUES	27
1.3.	PHAGOTHÉRAPIE	28
1.3.1.	PHAGES ET PHAGOTHÉRAPIE	29
1.3.1.1.	Généralités à propos des bactériophages.....	29
1.3.1.2.	Phagothérapie.....	31
1.3.2.	MODÉLISATION DE L'INTERACTION PHAGE/BACTÉRIE.....	33
1.3.2.1.	Modélisation de la cinétique bactérienne	34
1.3.2.2.	Modélisation de la dynamique virale.....	37
1.3.2.3.	Modèles mathématiques existants pour la description de l'interaction phage/bactérie..	39
1.4.	OBJECTIFS DE LA THÈSE	40

DEUXIÈME PARTIE : EVALUATION ET OPTIMISATION DE PROTOCOLES DANS LES MODÈLES MIXTES

..... 42

2.1.	PROTOCOLES ROBUSTES TENANT COMPTE DE L'INCERTITUDE DU MODÈLE POUR RÉPONSE BINAIRE RÉPÉTÉE	42
2.1.1.	RÉSUMÉ	42
2.1.2.	ARTICLE 1 (PUBLIÉ).....	44

2.2. ALGORITHMES POUR L'OPTIMISATION DE PROTOCOLES.....	68
2.2.1. ALGORITHMES MULTIPLICATIFS.....	68
2.2.1.1. Résumé	68
2.2.1.2. Article 2 (soumis)	69
2.2.2. CONTRIBUTION À LA NOUVELLE VERSION DE PFIM.....	121
2.3. COMPARAISON DE PROTOCOLES POUR ÉVALUER LA DÉCROISSANCE TUMORALE PRÉCOCE.....	122
2.3.1. RÉSUMÉ	122
2.3.2. ARTICLE 3 (ACCEPTÉ)	124
<u>TROISIÈME PARTIE : MODÉLISATION DE L'INTERACTION BACTÉRIOPHAGES/BACTÉRIES POUR LE TRAITEMENT DE LA PNEUMONIE</u>	<u>138</u>
3.1. RÉSUMÉ.....	138
3.2. ARTICLE 4 (SOUMIS)	140
<u>QUATRIÈME PARTIE : DISCUSSION</u>	<u>177</u>
4.1. PROTOCOLES EN MODÉLISATION	177
4.1.1. PROTOCOLES ROBUSTES PRENANT EN COMPTE L'INCERTITUDE DU MODÈLE	177
4.1.2. ALGORITHMES MULTIPLICATIFS.....	179
4.1.3. COMPARAISON DE PROTOCOLES POUR ÉVALUER LA DÉCROISSANCE TUMORALE PRÉCOCE	183
4.2. MODÉLISATION DE L'INTERACTION BACTÉRIOPHAGES/BACTÉRIES POUR LE TRAITEMENT DE LA PNEUMONIE ...	185
4.3. CONCLUSION GÉNÉRALE	188
<u>BIBLIOGRAPHIE.....</u>	<u>190</u>
<u>ANNEXE.....</u>	<u>203</u>

Première partie : Introduction

Les travaux de recherche composant ce doctorat ont reposé sur l'utilisation des modèles non-linéaires à effets mixtes (MNLEM), outils adaptés à l'analyse de données longitudinales. Ces travaux sont décomposés en deux parties. La première thématique concerne l'évaluation et l'optimisation de protocoles d'études longitudinales en amont de l'analyse par modélisation. Les travaux méthodologiques en planification d'essais constituent un champ de recherche important. Le choix du protocole est essentiel pour rendre les résultats d'analyse par MNLEM précis et donc permettre des inférences fiables. La seconde thématique est plus appliquée et utilise les MNLEM pour analyser des données précliniques de phagothérapie. La phagothérapie est une alternative à l'antibiothérapie pour le traitement des infections bactériennes, elle consiste en l'administration de bactériophages. Les bactériophages sont des virus infectant uniquement les bactéries. L'émergence de la résistance bactérienne aux antibiotiques est de plus en plus préoccupante, mener des études autour de la compréhension des mécanismes et des solutions pour parer à l'antibiorésistance est donc d'importance majeure. Il demeure nécessaire de bien comprendre et de quantifier les interactions entre bactéries et bactériophages.

Dans la première partie de cette introduction, nous décrivons les MNLEM. Les méthodes actuellement utilisées pour l'évaluation et l'optimisation de protocoles pour des essais analysés par MNLEM sont décrites dans la deuxième partie. Ensuite, la phagothérapie est présentée dans la troisième partie. Dans cette troisième partie, il est aussi présenté l'état de l'art des principaux modèles existants en bactériologie et en virologie qu'il est nécessaire de connaître pour la modélisation de données d'études précliniques sur la phagothérapie. Les objectifs de la thèse sont présentés en fin d'introduction.

1.1. Modèles non-linéaires à effets mixtes

Les modèles mathématiques visent à représenter les données issues du monde réel (Witelski and Bowen 2015). Ils peuvent aussi expliquer les mécanismes aboutissant aux observations ou être utilisés dans un objectif de prédiction. Ils peuvent être appliqués dans divers domaines scientifiques (physique, santé, sciences sociales, *etc.*) et se traduisent sous forme d'équations. Ces équations peuvent être sous forme de solutions analytiques. Elles peuvent aussi être sous la forme d'équations différentielles, notamment utilisées dans les systèmes dynamiques qui permettent de décrire l'évolution temporelle de différentes variables observées. Ce dernier type

de modèle est souvent utilisé dans le domaine de la santé, par exemple pour décrire l'évolution d'une épidémie dans une population ou encore pour décrire l'effet d'un médicament sur la progression d'une maladie à l'échelle individuelle. Souvent, ils visent à représenter les phénomènes le plus mécanistiquement possible, c'est à dire décrivant au mieux les processus biologiques sous-jacents. Le défi dans ce type d'approche est d'estimer les paramètres du modèle, compte tenu des données dont on dispose. En effet, un modèle plus complexe et donc avec davantage de paramètres peut généralement mieux caractériser les phénomènes réels. Néanmoins, si les informations apportées par les données ne sont pas suffisantes, l'estimation des paramètres est imprécise et il est fréquent de ne pas pouvoir identifier ces paramètres. Lors du choix du modèle et de sa paramétrisation, il est donc nécessaire de trouver un compromis entre la représentation fidèle des phénomènes biologiques qui sont complexes et la simplification de ces phénomènes, en fonction des connaissances et données à disposition.

Les modèles mixtes sont particulièrement adaptés à l'analyse de données longitudinales, c'est-à-dire les données recueillies à différents temps lors du suivi de chaque individu d'une population (Pinheiro and Bates 2000). Ce type d'approche permet de compenser le manque d'information pour un individu, en utilisant l'ensemble des données de tous les individus d'une étude. Les observations de tous les individus sont classiquement décrites par un même modèle mathématique structural. Les modèles mixtes permettent d'estimer, d'une part, les caractéristiques de population c'est-à-dire les paramètres moyens (effets fixes), et d'autre part la variabilité inter- voire intra-individuelle (effets aléatoires), ce qui permet d'obtenir des paramètres pouvant varier d'un individu à un autre et d'une occasion à une autre, respectivement. La variabilité peut aussi être « expliquée » par des covariables catégorielles (*e.g.* le génotype) ou continues (*e.g.* l'âge). Enfin la différence entre une observation et la prédiction du modèle est l'erreur résiduelle, c'est-à-dire la part non expliquée par le modèle. Cette erreur résiduelle peut comporter, entre autres, l'erreur de mesure et un mauvais choix de modèle.

Les MNLEM (Davidian and Giltinan 2017) sont fréquemment appliqués en pharmacométrie, l'étude de la pharmacologie quantitative qui comprend la pharmacocinétique (PK) et la pharmacodynamie (PD) (Ette and Williams 2007). La PK qui étudie le devenir d'une substance active contenue dans un médicament après son administration dans l'organisme, c'est-à-dire la concentration (le plus souvent plasmatique) d'un médicament au cours du temps. Les données de PK ont été analysées par MNLEM à partir des années 1970 (Sheiner et al. 1972). Les modèles permettent aussi de décrire la PD, c'est-à-dire l'effet du médicament sur l'organisme. Il est ainsi

possible de mettre en relation la concentration d'un médicament, son effet et le temps par l'utilisation d'un modèle PKPD. La PD peut alors être dépendante de la concentration courante de médicament (réponse directe), un délai peut aussi être pris en compte entre la concentration et l'effet (réponse indirecte). En outre, la PD peut être dépendante de l'exposition au médicament (mesurée par l'aire sous la courbe ou AUC) ou encore du pic de concentration. La pharmacométrie est aujourd'hui employée au cours des différentes étapes du développement d'un médicament, et dans toutes les aires thérapeutiques (Marshall et al. 2019).

Les MNLEM ont de nombreuses applications en pharmacométrie et sont de plus en plus employés pour aider à la prise de décision suite aux études cliniques (Kimko and Pinheiro 2015). De manière générale, le fait de prendre en compte l'ensemble des données longitudinales conduit à une plus grande puissance de tests. Premièrement, ils permettent de mettre en évidence des effets de covariables, par test de Wald ou de rapport de vraisemblance, et ces effets de covariables peuvent avoir un impact clinique. Lorsque des effets de covariables sont mis en évidence, cela peut conduire à des adaptations de posologie en fonction des caractéristiques d'un patient avant même le début de la thérapie, de manière à éviter des doses de médicament inefficaces ou, au contraire, toxiques. Par exemple, les effets de 4 covariables (âge, poids, sexe et créatininémie, marqueur de la fonction rénale) ont été mis en évidence sur la clairance du carboplatine, un anticancéreux (Chatelut et al. 1995). Une fois la thérapie débutée, l'adaptation de la posologie individuelle peut être réalisée. A partir d'un modèle et de quelques mesures chez un patient, il est possible d'estimer ses paramètres individuels, généralement obtenus à partir des EBE (*empirical Bayes estimates*), et ainsi définir la posologie adaptée pour atteindre la concentration, l'exposition voire l'effet cible. L'intérêt de cette méthode d'ajustement de posologie a par exemple été montré pour la vancomycine, un antibiotique (Berthaud et al. 2019). Une autre utilisation possible des modèles mixtes est la réalisation de tests statistiques basés sur les prédictions individuelles, c'est-à-dire à partir des EBE. A partir d'un MNLEM, il est aussi possible d'effectuer des simulations étudiant des conditions différentes de l'étude à partir de laquelle le modèle a été construit. Par exemple, dans le cadre de la maladie à virus Ebola, des posologies (doses de charge et d'entretien) pour de futures études ont été recommandées par simulations (Madelain et al. 2017). Il a aussi été montré l'impact du moment de l'initiation du traitement sur le risque de décès (Madelain et al. 2018). De manière plus générale, les MNLEM sont souvent utilisés pour la sélection de dose durant le développement pharmaceutique (Bonate 2000; Miller 2005). Enfin, les MNLEM sont utiles pour évaluer et optimiser les protocoles d'études (section 2.).

1.1.1. Modèles pour données continues

Les MNLEM, notamment en pharmacométrie, sont le plus souvent employés pour décrire des données continues, par exemple la concentration d'un médicament, la glycémie, la taille tumorale ou encore la charge bactérienne.

Pour décrire un vecteur de données continues y_i (de longueur n_i) d'un individu i ($i = 1, \dots, N$), un MNLEM peut s'écrire de la manière suivante (équation 1) :

$$y_i = f(\theta_i, \xi_i) + \varepsilon_i, \quad (1)$$

où la fonction f est le modèle structurel, θ_i est le vecteur des paramètres individuels, ξ_i est le protocole élémentaire (temps de mesure, schéma d'administration, *etc.*, voir section 1.2.) et ε_i le vecteur d'erreurs résiduelles individuel ($\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T$).

Dans ce type de modèle, f est donc une fonction mathématique commune à l'ensemble des individus. Les différences entre les individus sont décrites par les paramètres individuels θ_i , qui peuvent être donnés par la fonction $g(\mu, b_i, z_i, \beta)$ avec μ le vecteur de paramètres à effets fixes, b_i le vecteur d'effets aléatoires, z_i le vecteur des éventuelles covariables et β le vecteur d'effets de covariables. Supposons d'abord un modèle simple sans effet de covariable. La fonction g est le plus fréquemment additive ($\theta_i = \mu + b_i$) ou exponentielle ($\theta_i = \mu \times e^{b_i}$). Le vecteur d'effets aléatoires b_i suit une loi normale d'espérance 0 et de matrice de variance-covariance Ω . Lorsque g est additive, la distribution de θ_i est donc normale, lorsque g est exponentielle, la distribution de θ_i est log-normale. D'autres distributions sont possibles pour θ_i (logit-normale, probit-normale, multimodale, *etc.*) mais ne seront pas détaillées ici. Les effets aléatoires peuvent caractériser la variabilité inter-individuelle (BSV) ou la variabilité inter-occasion ou intra-individuelle (WSV), c'est-à-dire la variation des paramètres entre différentes périodes (Karlsson and Sheiner 1993).

Les effets de covariables peuvent expliquer une partie de la variabilité observée. Il existe deux grands types de covariables : catégorielle (par exemple le sexe, un variant génétique ou l'échelle de performance ECOG (*Eastern Cooperative Oncology Group*) à l'inclusion) ou continues (par exemple l'âge ou la masse corporelle). Pour une covariable catégorielle, il est commun de considérer une classe de référence et d'estimer l'effet des autres classes sur un paramètre. Les paramètres individuels peuvent ainsi s'écrire $\theta_i = \mu + b_i + \beta z_i$, avec $z_i = 0$ pour la classe de référence et $z_i = 1$ pour l'autre classe, pour une covariable binaire. Il est aussi possible de rassembler des individus de différentes classes afin de gagner en puissance statistique. Un

modèle de covariable peut s'écrire de différentes manières, notamment pour les covariables continues. En plus des effets de covariables à l'inclusion, l'effet de covariables temps-dépendantes peut être étudié, par exemple, la variation de la créatininémie d'un individu depuis le début de l'étude (Wählby et al. 2004).

Le vecteur de résidus est noté ε_i et est supposé suivre une loi normale de moyenne 0 et de variance Σ_i , avec Σ_i qui est souvent une matrice diagonale de taille n_i , définie telle que dans l'équation 2 :

$$\Sigma_i = \text{diag} \left(\sigma_{inter} + \sigma_{slope} \times f(\theta_i, \xi_i) \right)^2, \quad (2)$$

où σ_{inter} et σ_{slope} sont respectivement les paramètres d'erreur additive et proportionnelle. Des modèles d'erreur plus complexes peuvent aussi être utilisés.

Pour les modèles de données continues, le vecteur de paramètres estimés ψ , de longueur P , est composé de chaque effet fixe μ , de chaque effet de covariable β , des éléments de la matrice de variance-covariance Ω et des paramètres d'erreur σ_{inter} et σ_{slope} .

1.1.2. Modèles pour données non continues

De plus en plus, les MNLEM sont utilisés pour analyser des données longitudinales non continues. C'est-à-dire les données binaires, scores cliniques, données de comptage, temps d'événement, *etc.*

Les données binaires répétées analysées par MNLEM sont au centre de l'un des travaux de ce doctorat. Dans différents domaines, la réponse peut être définie de manière binaire. Les critères peuvent être de type oui/non pour définir, entre autres, la guérison, la présence d'un symptôme ou d'un effet indésirable. Par exemple, lors d'un test diagnostique du virus de l'immunodéficience humaine, la charge virale est détectable si supérieure à 200 copies par mL ou indétectable sinon (Rodger et al. 2019). Le portage bactérien peut être dichotomisé de manière similaire et mesuré à différents temps. Par exemple, dans une étude sur *Streptococcus pneumoniae*, des écouvillons ont été réalisés chez des enfants à différents temps, et la relation entre la probabilité de portage et différentes variables telles que l'âge ont été étudiées (Syrjänen et al. 2001). Les études sur la maladie à SARS-CoV2 2019 rapportent aussi des critères binaires pour le diagnostic, la rémission ou la séroconversion, bien que la dichotomisation de l'information en positif/négatif peut conduire à une perte de puissance statistique (Blain et al. 2020; Fragkou et al. 2020). Lorsque les données binaires sont recueillies de manière

longitudinale, leur analyse nécessite l'utilisation de modèles mixtes linéaires généralisés tels que les modèles logistiques qui peuvent s'écrire, pour une donnée binaire y_{ij} pour l'individu i et l'observation j , au temps de mesure t_{ij} (équation 3) :

$$\text{logit} \left(p(y_{ij} = 1 | b_i) \right) = f(\theta_i, t_{ij}). \quad (3)$$

Par exemple, pour un lien logit-linéaire avec le temps, la probabilité d'observer la modalité $y_{ij} = 1$ (équation 4) s'écrit :

$$p(y_{ij} = 1 | b_i) = \frac{\exp(\theta_{1i} + \theta_{2i} \times t_{ij})}{1 + \exp(\theta_{1i} + \theta_{2i} \times t_{ij})}, \quad (4)$$

où θ_{1i} et θ_{2i} sont respectivement les paramètres d'origine et de pente de l'individu i .

D'autres liens logit peuvent être utilisés (logarithmique, quadratique, exponentiel, *etc.*). Comme pour les modèles de données continues, les paramètres individuels θ_i sont donnés par la fonction $g(\mu, b_i, z_i, \beta)$ avec μ le vecteur de paramètres à effets fixes, b_i le vecteur d'effets aléatoires, z_i le vecteur des éventuelles covariables et β le vecteur d'effets de covariables.

Pour les modèles de données binaires répétées, le vecteur de paramètres estimés ψ , de longueur P , est composé de chaque effet fixe μ , de chaque effet de covariable β et des éléments de la matrice de variance-covariance Ω .

Au cours de cette thèse, des données de survenue d'événement ont aussi été analysées. Différents types d'analyse de ce type de données existent, les plus connues étant l'estimation non paramétrique selon Kaplan et Meier (Kaplan and Meier 1958) et semi-paramétrique selon Cox (Cox 1972). Dans les modèles de Cox (modèles à risques proportionnels), le risque instantané $h_i(t)$ pour l'individu i au temps t , qui correspond à la probabilité de survenue d'un événement sur un intervalle de temps sachant que l'événement ne s'est pas encore produit, s'écrit (équation 5) :

$$h_i(t|z_i) = h_0(t) \exp(\beta \times z_i), \quad (5)$$

avec $h_0(t)$ le risque de base et β l'effet de la covariable z_i .

La probabilité de survie de l'individu i $S_i(t)$ jusqu'à t est alors décrite par l'équation 6 :

$$S_i(t|z_i) = e^{-\int_0^t h_i(s|z_i) ds}. \quad (6)$$

Il est aussi possible d'étudier la survenue d'un événement par approche paramétrique, avec h_0 une fonction de risque de base dépendant de paramètres. Enfin, le lien entre un biomarqueur longitudinal et le risque instantané peut être établi. Lorsque cet événement est le décès, ce type

d'analyse s'appelle la modélisation conjointe (Tsiatis et al. 1995; Wu et al. 2012). Dans ce type d'analyse, les données longitudinales du biomarqueur sont décrites par un modèle, par exemple, un MNLEM, et la fonction de risque instantané peut s'écrire (équation 7) :

$$h_i(t|F_i(\theta_i, t), \theta_s) = h_0(t) \exp(\beta \times f(\theta_i, t)), \quad (7)$$

avec $F_i(\theta_i, t)$ l'historique de la prédiction individuelle jusqu'au temps t , θ_s les paramètres du modèle de survie (pour le risque de base et le lien avec le marqueur longitudinal), $h_0(t)$ la fonction décrivant le risque de base et $f(\theta_i, t)$ la prédiction individuelle de la variable longitudinale au temps t . D'autres fonctions de lien entre le risque instantané de décès et la variable longitudinale sont possibles, par exemple un lien non linéaire (Madelain et al. 2018). Pour les modèles conjoints, le vecteur de paramètres estimés ψ , de longueur P , est composé des paramètres du modèle longitudinal (*i.e.* pour données continues, voir section 1.1.1.) et des paramètres de survie.

Les modèles conjoints présentent certains avantages. Ils permettent de prédire la probabilité de survie future d'un individu de manière dynamique, à partir du modèle et de la trajectoire du marqueur longitudinal d'intérêt chez cet individu (Desmée et al. 2015; Riglet et al. 2020). Il a par exemple été montré que l'utilisation des données longitudinales dans un modèle en oncologie permet de meilleures prédictions que celle des données à l'inclusion uniquement (Tardivon et al. 2019). En outre, l'estimation des paramètres longitudinaux peut être biaisée lorsqu'un nombre conséquent d'individus sortent rapidement de l'étude de manière non aléatoire (décès ou à échec de la thérapie par exemple). Dans ce cas, pour un modèle qui ne tient pas compte de la survie (donc de la sortie d'étude), les valeurs de paramètres peuvent être influencées par les données des individus survivants, qui ne sont pas nécessairement représentatifs de l'ensemble de la population étudiée. Les modèles conjoints permettent de corriger ce biais (Rizopoulos 2012; Desmée et al. 2015).

1.1.3. Méthodes et logiciels d'estimation

La majorité des méthodes d'estimation de paramètres dans les MNLEM sont fondées sur le maximum de vraisemblance. Cette vraisemblance, dans le cas de modèles non linéaires, n'a pas de solution analytique. La vraisemblance L des observations y_i d'un individu i s'écrit (équation 8) :

$$L(y_i|\psi) = \int_{b_i} p(y_i|b_i, \psi)p(b_i|\psi)db_i, \quad (8)$$

avec $p(y_i|b_i, \psi)$ la distribution de probabilité conditionnelle (PDF) des y_i sachant b_i et $p(b_i|\psi)$ la PDF de b_i . Il existe différents algorithmes permettant d'estimer les paramètres des MNLEM. Le premier algorithme a été développé au début des années 1970 par Sheiner. Cet algorithme, nommé *First Order* (FO) repose sur une linéarisation du modèle autour de l'espérance des effets aléatoires (Sheiner et al. 1972). Des extensions de cette méthode ont été proposées ensuite : par exemple, l'algorithme *First Order Conditional Estimation* (FOCE) se base sur la linéarisation du modèle autour des estimations individuelles des effets aléatoires (Lindstrom and Bates 1990). Ces deux méthodes sont intégrées au logiciel NONMEM (<https://www.iconplc.com/innovation/nonmem/>, Bauer 2019a, 2019b). Une autre méthode populaire pour l'estimation des paramètres des MNLEM est l'algorithme *Stochastic Approximation Expectation Maximisation* (SAEM) couplé aux chaînes de Markov de Monte carlo (MCMC) (Kuhn and Lavielle 2005). A chaque itération, une première phase (« Expectation ») consiste à simuler des paramètres individuels par MCMC puis à estimer la vraisemblance. La seconde phase (« Maximisation ») consiste à déterminer les paramètres maximisant la vraisemblance. Les paramètres sont alors mis à jour pour l'itération suivante jusqu'à convergence. Cet algorithme est intégré aux logiciels NONMEM, MONOLIX (<http://lixoft.com/products/monolix/>) et au package R *saemix* (Comets et al. 2017). L'algorithme SAEM est généralement plus efficace que FOCE dans le cas de données éparées et de modèles complexes (Bauer 2019b).

Dans le cadre de ces travaux de thèse, les paramètres des NLMEM ont été estimés avec l'algorithme SAEM du logiciel MONOLIX. Cet algorithme permet de fournir des estimations pour des modèles décrivant différents types de données : continues, discrètes ou de survie.

1.1.4. Sélection et évaluation de modèles

Les méthodes de sélection de modèle sont une thématique majeure en biostatistiques et pharmacométrie. Usuellement en pharmacométrie, la comparaison de modèles se fait selon le critère d'Akaike (AIC) ou de Schwartz (BIC) qui sont fondés sur la vraisemblance (Delattre et al. 2014). Un modèle avec un AIC ou BIC plus faible est plus parcimonieux. La précision d'estimation des paramètres doit aussi être vérifiée. L'évaluation de la précision d'estimation des paramètres se fait en analysant l'erreur standard relative (RSE). Les RSE des paramètres sont obtenues en divisant l'erreur standard (SE) par l'estimation du paramètre, avec les SE des paramètres obtenues à partir de la racine carrée des éléments diagonaux de l'inverse de la matrice d'information de Fisher. Bien que ce critère ne soit pas clairement établi dans la

littérature, on considère qu'un paramètre est précisément estimé si la RSE ne dépasse pas 30 % pour les paramètres à effets fixes et 50 % pour les variances des effets aléatoires.

Au cours de la construction d'un modèle et avant d'utiliser un modèle, il est nécessaire de l'évaluer selon différents critères tels que l'adéquation entre données et prédictions du modèle et, si possible, la vérification des hypothèses émises. Récemment, des standards ont été proposés concernant l'évaluation de modèles, en particulier au moyen d'outils graphiques. Un tutoriel a été proposé par la société internationale de la pharmacométrie (ISoP) pour l'évaluation de modèles en pharmacométrie pour les données continues (Nguyen et al. 2017). Des recommandations ont aussi été émises par la FDA (*Food and Drug Administration*), l'agence réglementaire américaine, pour l'évaluation de modèles pharmacocinétiques de population (Food and Drug Administration 2019). Les principaux graphiques d'évaluation sont les observations en fonction des prédictions (individuelles ou moyennes), qui doivent être également répartis de part et d'autre de la droite d'identité, et les résidus (individuels, de population, pondérés, etc.) en fonction du temps ou des prédictions qui doivent être répartis de part et d'autre de l'axe $y = 0$. Les graphiques individuels, c'est-à-dire la représentation des prédictions et des observations en fonction du temps pour chaque individu, peuvent aussi être utilisés à des fins diagnostiques. Pour une évaluation avancée du modèle, des méthodes utilisant des simulations sont nécessaires. Les vérifications visuelles des prédictions (VPC, *Visual Predictive Check*) permettent de comparer les simulations basées sur le modèle aux observations pour différents percentiles (Yano et al. 2001). Les résidus simulés normalisés décorrélés (NPDE, *Normalized Prediction Distribution Error*) en fonction du temps ou des prédictions, doivent être répartis également autour de l'axe $y = 0$ et leur distribution doit être proche d'une loi Normale d'espérance nulle et de variance 1 (Brendel et al. 2006).

Bien qu'il n'y ait pas de standard concernant l'évaluation de modèles pour données non continues répétées à ce jour, quelques outils sont disponibles. Par exemple, pour les modèles de données binaires, des résidus fondés sur une variable latente ont été proposés (Liu and Zhang 2018). Pour les modèles de données de temps d'événement tels que les modèles conjoints, il est possible de procéder à l'évaluation du modèle par VPC construits autour de l'estimateur de la survie de Kaplan-Meier (Friberg et al. 2009; Björnsson and Simonsson 2011) ou encore par NPDE (Cerou et al. 2018).

1.1.5. Modélisation et oncologie

Le choix de protocoles des essais cliniques en oncologie est un point critique, en particulier pour les phases précoces, où la réponse, la toxicité et la PK des anticancéreux sont évaluées. Dans ces études, il est de plus en plus recommandé d'utiliser des biomarqueurs plutôt que des critères tels que la survie des patients qui nécessitent des études longues (Seymour et al. 2013). Parmi ces biomarqueurs, la taille tumorale est un marqueur d'efficacité intéressant. Particulièrement, la décroissance précoce de la taille tumorale (ETS, *Early Tumor Shrinkage*), mesurable après quelques semaines de traitement, est significativement liée aux critères plus cliniques, couramment utilisés en oncologie tels que la survie ou la survie sans progression (Claret et al. 2009; Wang et al. 2009).

La littérature est particulièrement riche concernant les modèles d'évolution de la taille tumorale, la plupart sont décrits dans une revue (Ribba et al. 2014) et disponibles dans les bibliothèques de modèles telles que DDMoRe (<http://repository.ddmore.eu/>). La taille tumorale est souvent mesurée comme la somme des plus longs diamètres des différentes lésions tumorales. Ces mesures sont néanmoins sujettes à une erreur de mesure potentiellement importante due à la variabilité entre les différents appareils d'imagerie et les opérateurs (Sullivan et al. 2013). Cela renforce l'utilité des MNLEM, qui prennent en compte ces erreurs de mesure au travers de l'estimation de l'erreur résiduelle.

Parmi les modèles les plus anciens, le modèle de Gompertz (équation 9) est utilisé pour décrire l'évolution naturelle de la taille tumorale en l'absence de traitement, c'est-à-dire une croissance exponentielle au début puis un ralentissement de la vitesse de croissance jusqu'à un plateau (Gompertz 1825). Il peut s'écrire sous cette forme :

$$\frac{dT_S}{dt} = KS \times T_S \times \ln \left(\frac{T_S}{T_{S_{max}}} \right), \quad (9)$$

avec T_S la taille tumorale, t le temps, KS le paramètre de croissance tumorale et $T_{S_{max}}$ la taille tumorale maximale.

Simeoni a développé un modèle à partir de données chez la souris qui prend en compte les différents états cellulaires, et utilise donc des compartiments de transduction où les cellules tumorales prolifèrent de manière exponentielle et s'engagent vers les phases de mort cellulaires, notamment lorsqu'exposées à des traitements anticancéreux (Simeoni et al. 2004). Les modèles d'inhibition de la croissance tumorale (TGI, *Tumor Growth Inhibition*) peuvent prendre différentes formes : la croissance peut être linéaire, exponentielle, de type Gompertz ou autre

et la décroissance médiée par la concentration ou l'exposition médicamenteuse, avec ou sans résistance au traitement. Ces modèles peuvent s'appliquer à différents types de cancer et de thérapies. Par exemple, le modèle de Claret (équation 10) s'écrit :

$$\frac{dT_S}{dt} = KS \times TS - KD \times AUC_X \times e^{-KR \times t} \times TS, \quad (10)$$

avec TS la taille tumorale, t le temps, KS le paramètre de croissance tumorale, KD la décroissance tumorale liée à l'exposition (AUC_X) à l'anticancéreux X et KR la résistance à X . Ce modèle a été développé dans le cadre d'une étude sur le cancer colorectal (Claret et al. 2009) mais a aussi été utilisé pour des données d'étude sur le cancer thyroïdien (Lu et al. 2010).

La modélisation est aussi utile en phase 2 ou en phase 3, lorsque l'efficacité de la thérapie est souvent évaluée sur la survie ou la survie sans progression. En particulier, la méthodologie des modèles conjoints constitue un outil idéal pour l'analyse de ces données d'événement, en liant la pharmacocinétique du ou des médicaments, voire l'évolution d'un biomarqueur (par exemple, la taille tumorale (Tardivon et al. 2019) ou le PSA (*Prostate Specific Antigen*) dans le cancer de la prostate (Desmée et al. 2015)) au risque de décès.

En oncologie, des combinaisons de médicaments sont fréquemment l'objet d'essais cliniques. Lors de l'utilisation de MNLEM dans ce cadre, il est préférable de quantifier l'effet propre de chaque molécule impliquée ainsi que l'interaction entre les différentes molécules : antagonisme, additivité ou synergie, par exemple par le biais d'un modèle d'interaction pharmacodynamique globale (Wicha et al. 2017). Néanmoins, afin d'identifier tous ces éléments, il est nécessaire de disposer de données informatives, ce qui est possible si le protocole de l'étude a été élaboré judicieusement en amont.

1.2. Evaluation et optimisation de protocoles dans les modèles mixtes

Les modèles mixtes permettent d'analyser des données éparses, c'est-à-dire avec peu d'information par individu (nombre réduit de temps de mesures par exemple) par rapport au nombre de paramètres du modèle que l'on souhaite estimer. En amont de la réalisation d'une étude, son protocole doit être défini. En optimisation de protocoles, on s'intéresse aux méthodes permettant de récolter le maximum « d'information » dans le but de répondre aux objectifs d'une étude, en prenant en compte les contraintes imposées par celle-ci. Afin d'évaluer et d'optimiser un protocole avant analyse par MNLEM, les deux méthodes possibles sont

l'utilisation de la matrice d'information de Fisher (FIM) ou la simulation d'essais cliniques CTS (ou non-cliniques). Une première méthode d'évaluation de la FIM pour les modèles mixtes a été proposée à la fin des années 1990 (Mentré et al. 1997). Cette thématique de recherche a ensuite fait l'objet de nombreux développements (davantage détaillés en section 1.2.1.) : des travaux méthodologiques et appliqués; un congrès : le PODE (*Population Optimum Design of Experiments*, <http://www.maths.qmul.ac.uk/~bb/PODE/PODE.html>); des logiciels tels que PFIM, développé dans notre équipe de recherche.

Pour un individu i ($i = 1, \dots, N$), le design élémentaire, noté ξ_i , est défini par le nombre de mesures n_i , par les temps de mesures $(t_{i1}, \dots, t_{in_i})$ et possiblement par d'autres éléments telle que la dose d_i d'un médicament. Le design de population, noté \mathcal{E} , est défini par le nombre total d'individus N et par l'ensemble des designs élémentaires de chaque individu : $\mathcal{E} = \{\xi_1, \dots, \xi_N\}$. Les designs élémentaires sont souvent regroupés en Q groupes d'individus dans un design de population. Au sein d'un groupe q ($q = 1, \dots, Q$), tous les individus N_q partagent alors le même protocole élémentaire ξ_q . Le protocole de population peut alors être défini tel que $\mathcal{E} = \{[\xi_1, N_1], \dots, [\xi_Q, N_Q]\}$. Déterminer un protocole d'étude consiste à choisir le nombre d'individus, le nombre de groupes et d'individus par groupe, les temps de mesures, le nombre de temps de mesures ou la posologie (dose de charge, dose d'entretien, nombre de doses, intervalle de temps entre deux doses). Le choix du design influence directement la précision d'estimation des paramètres. En amont de la modélisation, le choix du design est donc crucial car mal choisi, il peut conduire à des études ininterprétables ou non concluantes (Jonsson et al. 1996). De plus, un bon choix de protocole peut aussi permettre de rendre une étude plus éthique et de réduire ses coûts, en ne mesurant la ou les variables d'intérêts qu'aux temps les plus informatifs, voire en réduisant le nombre d'individus au strict nécessaire pour répondre aux objectifs.

1.2.1. Matrice d'information de Fisher

Au début du 20^{ème} siècle, Fisher a montré le lien existant entre l'information obtenue à partir de données expérimentales et la matrice d'information de Fisher (Fisher 1922). L'approche basée sur l'évaluation de la FIM (Atkinson et al. 2007; Fedorov and Leonov 2013; Pronzato and Pázman 2013) est une alternative efficace à la méthode par CTS. Selon l'inégalité de Cramér-Rao, l'inverse de la FIM est la borne inférieure de la matrice de variance-covariance de tout estimateur non biaisé des paramètres. En d'autres termes, la racine carrée de l'inverse

des éléments diagonaux de la FIM constitue une prédiction des erreurs standard (SE) d'estimation des paramètres du modèle.

La FIM \mathcal{M} d'un protocole élémentaire ξ_i peut être exprimée de la manière suivante (équation 11) :

$$\mathcal{M}(\psi, \xi_i) = E_{y_i} \left(\frac{\partial \log(L(y_i|\psi))}{\partial \psi} \frac{\partial \log(L(y_i|\psi))}{\partial \psi}^T \right), \quad (11)$$

avec ψ le vecteur de paramètres à estimer, y_i le vecteur d'observation de l'individu i (dépendant notamment de ξ_i) et L la vraisemblance.

La FIM d'un protocole de population \mathcal{E} est en fait la somme des N FIM élémentaires (équation 12) :

$$\mathcal{M}(\psi, \mathcal{E}) = \sum_{i=1}^N \mathcal{M}(\psi, \xi_i). \quad (12)$$

Dans le cas particulier où tous les individus d'une étude partagent le même protocole élémentaire, la FIM de population est donc égale à $\mathcal{M}(\psi, \xi) \times N$.

La FIM est donc obtenue à partir du modèle et son vecteur de paramètres associés, pour un design de population. Il est ensuite possible de prédire la SE des paramètres d'un modèle à partir de la FIM. Cet outil permet donc d'éviter des protocoles conduisant à des estimations de paramètres trop peu précises pour être interprétées. Il est aussi possible de prédire la puissance du test de Wald pour détecter un effet de covariable catégorielle à partir de la FIM, la statistique du test de Wald étant $S_W(\beta) = \beta/SE(\beta)$ avec β l'effet de covariable estimé (Retout et al. 2007). Par extension, il est possible de prédire le nombre de sujets nécessaires pour atteindre une certaine puissance de ce test de Wald (classiquement 80 ou 90 %). De même, la puissance du test d'équivalence entre deux groupes peut être prédite à partir de la FIM (Nguyen et al. 2012).

En plus de l'estimation de la matrice individuelle et de population, la matrice d'information bayésienne pour les effets aléatoires peut être évaluée (Mentré et al. 1995). A partir de cette matrice bayésienne, il est possible de prédire la régression vers la moyenne des paramètres individuels ou η -shrinkage (Combes et al. 2013).

1.2.1.1. Méthodes d'évaluations de la FIM

Comme énoncé dans la section 1.1.1.3., la vraisemblance, et donc la FIM n'ont pas de forme analytique dans les MNLEM. Plusieurs méthodes ont été proposées pour évaluer la FIM. La méthode FO (*First Order*), première méthode qui a été proposée pour évaluer la FIM dans les MNLEM, utilise une linéarisation du premier ordre du modèle autour des effets aléatoires (Mentré et al. 1997). Initialement proposée dans le cadre d'un modèle d'erreur additif, cette méthode a été étendue à l'évaluation de la FIM pour MNLEM avec erreur proportionnelle (ou combinée) (Retout et al. 2002). Cette méthode est très efficace en général et intégrée à la plupart des logiciels d'optimisation de protocoles basés sur la FIM (Nyberg et al. 2015). Cependant, la FIM prédite, après calcul par FO, peut présenter certaines limites en cas de variabilités élevées, ou modèles très complexes, et particulièrement dans les modèles pour données discrètes. De nombreuses méthodes alternatives ont donc été étudiées pour évaluer la FIM pour ces modèles (Waite and Woods 2015). Des méthodes récentes ont été développées dans le but d'évaluer la FIM pour des modèles de données continues et discrètes. Parmi celles-ci, des méthodes sont fondées sur l'utilisation de simulations de Monte Carlo (MC), combinées à la quadrature de Gauss adaptative (AGQ) ou au Monte Carlo Hamiltonien (HMC).

L'approche de Monte Carlo-quadrature de Gauss adaptative MC-AGQ a d'abord montré de bonnes performances pour l'évaluation de la FIM dans le cadre de modèles présentant une non-linéarité importante (Nguyen and Mentré 2014), puis pour des modèles de données discrètes : binaires, binaires répétées, comptage, temps d'événements répétés (Ueckert and Mentré 2017). En termes de temps de calcul, cette méthode est efficace pour des modèles avec peu d'effets aléatoires mais présente un coût croissant avec le nombre d'effets aléatoires. Cette méthode a donc été étendue très récemment : l'approche QRMC-AGQ (*Quasi Random Monte Carlo-AGQ*) permet de réduire le temps de calcul tout en conservant une évaluation correcte de la FIM (Ueckert and Mentré 2015).

L'approche de Monte Carlo-Monte Carlo Hamiltonien MC-HMC a aussi montré de bons résultats pour les modèles complexes tels que les modèles de données discrètes, en conservant un temps de calcul acceptable avec un nombre important d'effets aléatoires (Riviere et al. 2016). Cette approche est codée en R dans la librairie *MIXFIM* (Riviere and Mentré 2015), et nécessite l'écriture de modèles en Stan, langage initialement développé pour les approches bayésiennes (Stan Development Team 2016).

1.2.1.2. Critères d'optimalité

Quelle que soit la méthode utilisée pour évaluer la FIM, il est nécessaire de comparer les différents protocoles évalués selon un critère. Lorsque l'on cherche à optimiser un protocole, il est possible de comparer directement les SE prédites ou la puissance prédite du test de Wald, mais il est plus commun d'utiliser des critères d'optimalité calculés à partir de la FIM. Le calcul de la FIM nécessite de spécifier la structure du modèle et de supposer ses valeurs de paramètres. Les connaissances sur le modèle et les paramètres sont généralement obtenues à partir de précédentes études. L'ensemble des critères définis évoqués dans cette section sont utilisés dans le cadre de planification de protocole locale, c'est-à-dire avec un modèle et ses valeurs de paramètres définis.

De manière générale, maximiser l'information de Fisher revient à minimiser la variance autour de l'estimation de paramètres, et donc à maximiser la précision d'estimation. Le critère de D-optimalité φ_D , largement utilisé, est directement calculé à partir du déterminant de la FIM (Kiefer 1959). Le protocole pour lequel le critère D est maximisé est celui pour lequel la précision d'estimation attendue est la meilleure en considérant l'ensemble des paramètres ψ , dont le vecteur est de dimension P . Le critère D est défini par l'équation 13 :

$$\varphi_D(\mathcal{E}) = \det(\mathcal{M}(\psi, \mathcal{E}))^{1/P}. \quad (13)$$

D'autres critères peuvent être utilisés (Atkinson et al. 2007). Par exemple, le critère A est calculé à partir de la trace de la FIM et le critère E à partir des valeurs propres. Le choix du critère d'optimalité peut dépendre de l'objectif de l'étude. Par exemple, lorsque l'on cherche à optimiser spécifiquement la précision d'estimation d'un paramètre ou d'un sous-ensemble de paramètres, le critère D_S (φ_{D_S}) peut être choisi (Karlin and Studden 1966). Le calcul de ce critère (équation 14), proche de celui du critère D, vise à maximiser l'information sur un sous-ensemble de paramètres d'intérêt (ψ_S , de longueur S) :

$$\varphi_{D_S}(\mathcal{E}) = \left(\frac{\det(\mathcal{M}(\psi, \mathcal{E}))}{\det(\mathcal{M}(\psi_T, \mathcal{E}))} \right)^{1/S}, \quad (14)$$

avec ψ composé du vecteur de paramètres d'intérêt ψ_S et du vecteur des autres paramètres ψ_T .

Ce choix de critère peut être intéressant dans diverses situations, par exemple si l'objectif d'une étude est de détecter un effet de covariable par test de Wald, la puissance de ce test étant dépendante de la précision d'estimation de cet effet (voir section 1.2.1.). Cependant, négliger la précision d'estimation des autres paramètres d'un modèle est déconseillé. Un design D_S -

optimal peut conduire à des situations où les autres paramètres du modèle ne sont pas identifiables (Sölkner 1993). Il est donc préférable d'utiliser des critères mélangeant le critère D_S à un autre critère. Par exemple, le critère DD_S (Atkinson and Bogacka 1997) combine les critères D et D_S , il s'écrit (équation 15) :

$$\varphi_{DD_S}(\mathcal{E}) = \det(\mathcal{M}(\psi_T, \mathcal{E}))^{\frac{1-\alpha}{p-s}} \times \left(\frac{\det(\mathcal{M}(\psi, \mathcal{E}))}{\det(\mathcal{M}(\psi_T, \mathcal{E}))} \right)^{\frac{\alpha}{s}}, \quad (15)$$

où α , compris entre 0 et 1, est un terme d'équilibre entre l'information sur les paramètres ψ_T et sur les paramètres d'intérêt. Plus récemment, le critère D_S , combiné à un autre critère (le critère V , utilisé dans le but de minimiser la variance des prédictions), a montré son utilité pour le choix de doses à évaluer dans le cadre d'une combinaison de deux molécules (Papathanasiou et al. 2019).

Afin de comparer facilement deux protocoles, il est commun de calculer l'efficacité pour un critère X , défini comme le ratio de ce critère pour le protocole \mathcal{E} par rapport à un protocole de référence \mathcal{E}^* , qui peut être le protocole optimal (équation 16) :

$$E_X(\mathcal{E}) = \frac{\varphi_X(\mathcal{E})}{\varphi_X(\mathcal{E}^*)}. \quad (16)$$

1.2.1.3. Incertitude du modèle et des paramètres

Dans le cas où le modèle final (c'est-à-dire celui effectivement utilisé pour décrire les données) et/ou les valeurs de paramètres sont différents du modèle spécifié *a priori* au moment du choix de protocole, le protocole choisi peut ne pas être assez informatif et conduire à des estimations peu précises de paramètres. Il est donc nécessaire de développer des approches de planification de protocole robuste, c'est-à-dire des méthodes prenant en compte l'incertitude sur le choix du modèle et des valeurs de paramètres. En cas d'incertitude, une première approche consiste à effectuer une analyse de sensibilité aux valeurs de paramètres, c'est-à-dire d'évaluer, voire optimiser un protocole avec différentes valeurs de paramètres. Des critères d'optimalité robustes, supposant une incertitude sur la valeur des paramètres, ont été proposés, par exemple le critère ED ou le critère HCD (Pronzato and Walter 1985; Foo and Duffull 2010). D'autres approches, plus récentes, ont été utilisées pour prendre en compte l'incertitude du modèle, dans le contexte où un protocole commun était choisi dans une étude d'un « cocktail » de médicaments (Nguyen et al. 2016). L'approche proposée dans ce travail était de supposer plusieurs modèles candidats et d'allouer un poids à chacun, sur le principe du *model averaging*

(Buckland et al. 1997; Buatois et al. 2018) et en utilisant la théorie de design optimal composé (Atkinson et al. 2007). Pour cette approche, différents modèles candidats doivent être définis et le critère robuste (ou composé) CD φ_{CD} (équation 17) peut être utilisé afin de trouver le protocole optimal commun aux modèles proposés. Le calcul de ce critère est en fait le produit pondéré des critères D de chaque modèle :

$$\varphi_{CD}(\mathcal{E}) = \prod_{m=1}^M \left(\varphi_{D,m}(\mathcal{E}) \right)^{w_m}, \quad (17)$$

avec m , un modèle candidat parmi les M modèles considérés et w_m le poids attribué à m (w_m tel que $0 \leq w_m \leq 1$ et $\sum_{m=1}^M w_m = 1$). Un w_m plus proche de 1 signifie que le modèle m pourrait mieux décrire les données que les autres modèles candidats.

Appliqué aux modèles de Poisson pour données de comptage, un travail d'optimisation de protocole robuste, basé sur la FIM calculée par MC-HMC, a été conduit de manière à prendre en compte l'incertitude sur les paramètres pour cinq modèles différents, puis sur les paramètres et considérant l'incertitude du modèle à choisir parmi les cinq (Loingeville et al. 2019). La prise en compte de l'incertitude des valeurs de paramètres n'a pas eu d'impact sur le protocole optimal dans la plupart des cas. Néanmoins, le fait de pré-spécifier un modèle qui n'est pas le modèle finalement utilisé peut conduire à de mauvais choix de protocoles, conduisant à des D-efficacités inférieure à 30% par rapport au protocole optimal (pour la même quantité de données recueillies). L'utilisation du critère CD a conduit à un choix de doses robuste, le protocole CD-optimal conduisant à une D-efficacité d'au moins 80% pour les 5 modèles candidats. Concernant les travaux de cette thèse, ce type d'approche robuste par rapport au choix du modèle a donc été retenue (section 2.1.).

1.2.1.4. Logiciels et algorithmes d'optimisation

Le premier logiciel qui a été développé pour optimiser des protocoles basés sur la FIM de population est PFIM (Retout et al. 2001; Dumont et al. 2018), dont la version 4.0. est disponible sur R et en version interfacée. D'autres logiciels de design ont ensuite été développés, par exemple POPT (Duffull et al. 2005), PopED (Foracchia et al. 2004), PopDes (Gueorguieva et al. 2007) et PkStaMp (Aliev et al. 2012). Les différents logiciels précédemment cités ont été comparés en 2014 et ont montré des résultats proches en termes de FIM prédite par méthode FO pour modèles de données continues (Nyberg et al. 2015). En outre, ce travail a aussi montré que l'utilisation de la matrice de Fisher diagonale par blocs est plus proche des résultats

empiriques (obtenus par simulations) que la matrice de Fisher complète. Parmi les spécificités de chaque logiciel au moment de cette comparaison, PopED permettait de prendre en compte l'incertitude sur la valeur des paramètres et PFIM permettait de prendre en compte la FIM observée d'une précédente étude (aujourd'hui, PopED a aussi cette fonctionnalité).

Les algorithmes développés pour l'optimisation de protocoles peuvent être séparés en deux catégories : ceux qui permettent de trouver un protocole dans un espace continu et ceux utilisant un espace de protocoles discret. Un algorithme utilisé pour une recherche de protocole optimal dans un espace continu est l'algorithme Simplex (ou du simplexe), développé à partir de 1947 pour des fonctions linéaires par Dantzig (Nash 2000) puis étendus à des problèmes d'optimisation non-linéaires (Nelder and Mead 1965). Le Simplex est l'algorithme intégré à PFIM 4.0. pour optimiser un protocole dans un espace continu. D'autres algorithmes ont été développés pour l'optimisation de protocoles pour MNLEM. Dans PopED, une combinaison de deux algorithmes (un permettant une recherche aléatoire globale et un autre, fondé sur le gradient stochastique, permettant d'affiner le résultat) est intégrée afin de chercher le protocole optimal dans un espace continu (Foracchia et al. 2004). Dans ce même but, l'utilisation du recuit simulé est une autre possibilité, il est notamment intégré à WinPOPT, la version interfacée de POPT (Duffull et al. 2002).

L'optimisation dans un espace discret consiste à déterminer le ou les designs élémentaires les plus informatifs à partir des possibles éléments de designs pré-définis tels que temps de mesure ou doses possibles. Dans PFIM 4.0., l'algorithme de Fedorov-Wynn (FW) est utilisé pour ce type d'optimisation (Fedorov 1972; Wynn 1972). Dans PopED, une version modifiée de cet algorithme est intégrée. L'algorithme FW permet aussi de définir la proportion d'individus qu'il faut allouer aux protocoles élémentaires. Simplex et FW, tous deux présents dans PFIM 4.0., ont été comparés. Il a été montré que FW est moins sensible que Simplex aux valeurs initiales et que FW est plus performant pour les problèmes d'optimisation complexes (Retout et al. 2007). D'autres algorithmes existants pourraient être adaptés à l'optimisation de protocoles d'études analysées par modèles mixtes. C'est l'objectif de l'un des travaux de cette thèse que de développer un algorithme pour optimiser ces protocoles (section 2.2.1.).

1.2.2. Simulations d'essais cliniques

La méthode par simulations consiste à estimer les paramètres d'un modèle à partir de différents jeux de données simulés à partir d'un modèle et du protocole à évaluer. Cette approche permet de juger de l'informativité d'un protocole. Comme pour l'approche basée sur la FIM, elle

permet d'évaluer la précision d'estimation des paramètres, en calculant l'écart type empirique des valeurs estimées. Cette approche permet aussi de vérifier la justesse d'estimation d'un paramètre ; un protocole informatif permet de minimiser le biais d'estimation des paramètres, calculé comme l'écart moyen entre valeur estimée et valeur réelle. Il est aussi souvent reporté l'erreur quadratique moyenne qui mesure à la fois la précision et le biais d'estimation.

Cette approche possède différents types d'applications. Au début des années 1990, sa première utilité était de comparer le biais et la précision des paramètres entre différents protocoles en termes de nombre et d'allocation des temps de mesure en pharmacocinétique (al-Banna et al. 1990). Cette approche a aussi rapidement servi à évaluer différents régimes de doses et d'évaluer l'intérêt d'une augmentation de doses en termes de pourcentage d'individus qui atteignent un seuil d'efficacité désiré (Green and Duffull 2003). De manière similaire, les EBE obtenus après simulations peuvent être utilisés pour réaliser des tests et comparer les puissances ou les erreurs de type 1 obtenues. Un des travaux de recherche de ce doctorat utilise cette approche dans ce but, avec une application en oncologie (section 2.3.). L'approche par CTS est aussi utile pour vérifier les prédictions issues de l'approche basée par la FIM. Cela est notamment utile pour comparer les logiciels (Nyberg et al. 2015) ou pour développer de nouvelles méthodes d'évaluation de la FIM (Nguyen and Mentré 2014).

L'approche par CTS est néanmoins peu adaptée à l'optimisation de protocoles. En effet, pour un protocole à évaluer, après avoir simulé un grand nombre de jeu de données, il est nécessaire d'estimer les paramètres du modèle à partir de chaque jeu de données, ce qui est coûteux en temps de calcul. Par conséquent, elle ne permet d'évaluer qu'un nombre limité de protocoles et le risque avec cette méthode est de passer à côté d'autres protocoles bien plus informatifs.

En pratique, lorsqu'on optimise un protocole, la meilleure approche semble donc de trouver le protocole le plus informatif par une approche basée sur la FIM. Cependant, comme expliqué dans la section 1.2.1., la FIM est basée sur la vraisemblance, elle-même calculée par approximation dans le cadre des MNLEM. De plus, l'inverse de la FIM donne la borne inférieure de la matrice de variance-covariance de tout estimateur non biaisé des paramètres. Il est donc conseillé de vérifier par CTS que le protocole optimal permet effectivement d'atteindre l'objectif souhaité.

1.3. Phagothérapie

Les antibiotiques ont drastiquement diminué la mortalité associée aux maladies infectieuses.

Leur utilisation massive et parfois inadaptée, en thérapeutique humaine ou chez les animaux d'élevage a conduit à l'émergence de bactéries résistantes à ces molécules. Le rapport O'Neill, établi en 2014 alertait sur l'augmentation de l'antibiorésistance et ses conséquences (O'Neill 2014). Il est estimé que la résistance bactérienne est responsable d'environ 700 000 décès par an dans le monde. Dans ce rapport, il est aussi prédit qu'en 2050, la résistance bactérienne pourrait être la première cause de mortalité avec 10 millions de décès par an. Dans ce contexte, la résistance bactérienne est particulièrement surveillée pour certaines espèces bactériennes, c'est le cas d'*Escherichia coli* par exemple. Il est aussi recommandé de promouvoir une meilleure hygiène et un usage plus rationnel des antibiotiques, c'est-à-dire utiliser les antibiotiques les plus adaptés, et seulement en cas de nécessité, que ce soit en thérapeutique humaine ou dans l'industrie agro-alimentaire (McEwen and Collignon 2018). De plus, la recherche de nouvelles thérapies est nécessaire, que ce soit le développement de nouveaux antibiotiques ou l'exploration d'autres voies telles que les anticorps dirigés contre les toxines bactériennes ou la phagothérapie : l'administration de virus dirigés contre les bactéries.

1.3.1. Phages et phagothérapie

Les bactériophages sont des virus prédateurs des bactéries. Ils ont été découverts en plusieurs étapes. Hankin a été le premier à mettre en évidence l'existence naturelle d'un agent antibactérien dans l'eau du Gange et de la Jumna à la fin du XIX^{ème} siècle (Hankin 1896). En 1917, d'Hérelle a isolé et démontré la spécificité de ces virus pour les bactéries qu'il a nommé « bactériophages », puis a proposé de les utiliser en thérapeutique (« phagothérapie ») contre les dysenteries à *Shigella* (d'Herelle 1917). Supplantés par les antibiotiques tels que la pénicilline à partir des années 1940, ils sont alors peu utilisés à l'échelle mondiale, hormis dans certains instituts d'Europe de l'Est où la recherche et le développement de la phagothérapie a perduré (Sulakvelidze and Morris 2001). Aujourd'hui, à cause de l'émergence de la résistance bactérienne aux antibiotiques, un regain d'intérêt pour la phagothérapie est observé dans les pays occidentaux, notamment en utilisation compassionnelle, pour des patients infectés par des bactéries toto-résistantes, c'est-à-dire résistantes à tout antibiotique sur le marché (Patey et al. 2018).

1.3.1.1. Généralités à propos des bactériophages

Les bactériophages infectent spécifiquement les bactéries. Il existe quatre types de cycles d'infection des bactéries par les phages (Figure 1). Le plus intéressant en thérapeutique est le

cycle lytique : après adsorption du phage à la surface de la bactérie, l'acide nucléique est transféré dans le cytoplasme. La transcription et la traduction de gènes se fait alors dans le but de synthétiser du matériel protéique et génomique viral plutôt que bactérien, afin de produire de nouveaux bactériophages. Vient ensuite la lyse bactérienne donc la libération des nouveaux virions qui infecteront d'autres bactéries. Parmi les autres cycles d'infection, la lysogénie et pseudolysogénie consistent en une possible lyse bactérienne après une ou plusieurs divisions bactériennes. Dans le cycle chronique, les bactéries deviennent productrices de phages sans lyse (Weinbauer 2004).

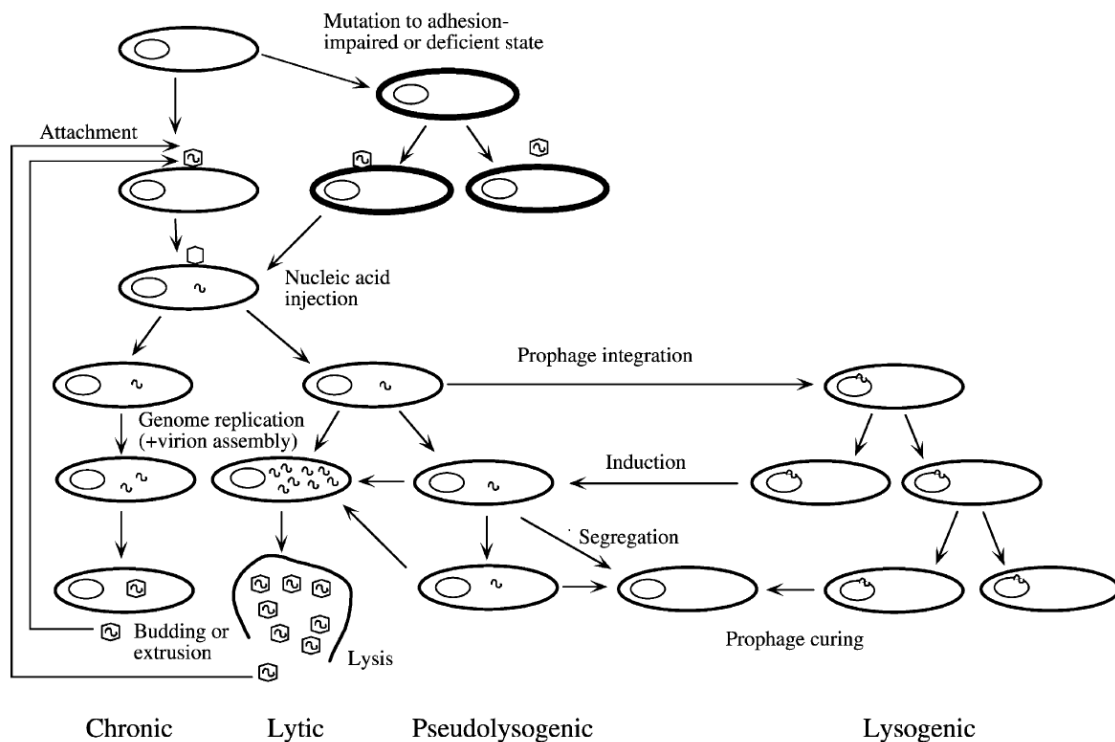


Figure 1. Les différents cycles d'infection des phages (d'après Weinbauer 2004).

Les bactériophages sont présents dans tous les écosystèmes, notamment dans notre alimentation et donc notre microbiote intestinal (Letarov and Kulikov 2009). Il est estimé que le nombre de phages sur Terre est de 10^{30} , soit en quantité plus importante que les bactéries. La diversité génétique des phages est aussi très importante, avec 10^8 génomes différents (Rohwer 2003). Les bactériophages participent de manière importante à l'équilibre de la biomasse bactérienne, via une relation proie (les phages) – prédateur (les bactéries), on parle aussi de couples phage bactérie par exemple, le couple 536/536_P1 où 536 est la souche bactérienne *Escherichia coli* 536 et 536_P1 un phage particulièrement virulent vis-à-vis de cette souche. Comme dans tout système proie-prédateur, des mécanismes sont mis en jeu par les bactéries pour résister aux phages, et par les phages pour contrer ces résistances. Par ailleurs, il a été démontré que les

bactériophages ont une part de responsabilité dans l'extinction des épidémies bactériennes (Faruque et al. 2005).

1.3.1.2. Phagothérapie

La préparation de phages avant administration chez l'animal ou l'homme nécessite le recueil du ou des phages d'intérêt, leur amplification et leur purification. Généralement, le spectre d'action d'un bactériophage est réduit. Il est souvent spécifique d'une espèce bactérienne et d'un nombre limité de souches au sein de cette espèce. Certains phages sont tout de même capable d'infecter différentes espèces (Jensen et al. 1998).

Cette particularité a un impact en thérapeutique. Lorsque l'infection est chronique avec un pronostic vital peu engagé, ou lorsque la bactérie responsable est identifiée, la phagothérapie peut être intéressante. Elle l'est d'autant plus si le lieu de l'infection est un tissu où les antibiotiques diffusent mal car les phages, contrairement aux médicaments chimiques, vont pouvoir se multiplier une fois sur le lieu de l'infection. De plus, leur spectre généralement étroit permet d'éviter la destruction des microbiotes (notamment intestinal) comme lors du traitement par antibiotiques. Dans le cas d'une infection aiguë avec pronostic vital engagé à court terme, lorsque la bactérie n'est pas identifiée ou lorsque l'infection est causée par plusieurs espèces bactériennes, il semble préférable, en première intention, d'utiliser des antibiotiques à large spectre (Wunderink and Waterer 2014).

De nombreuses preuves d'efficacité de la phagothérapie existent dans la littérature, notamment chez l'animal, tant en termes de réduction de la charge bactérienne que de mortalité (Caflisch et al. 2019). Cela a été montré pour différentes bactéries et pour divers types d'infection. Chez le lapin infecté par *Vibrio cholerae*, responsable du choléra, une réduction importante de la charge bactérienne et une absence de symptômes a été constatée pour les animaux traités par phages, contrairement aux non traités (Bhandare et al. 2019). Dans un projet précédent conduit dans le cadre d'une collaboration entre l'Institut Pasteur et notre unité de recherche, chez un modèle murin de pneumonie à *E.coli* 536, l'administration d'une dose de phage 536_P1 a fait aussi bien que 5 doses d'antibiotiques en termes de réduction de la charge bactérienne et de mortalité (Dufour et al. 2015). L'administration de plusieurs espèces de bactériophages permet d'étendre le spectre d'action de cette thérapie. Par exemple, l'étude d'un cocktail de trois phages chez la souris infectée par *K.pneumoniae* a montré un meilleur état de santé des souris traitées par le cocktail comparativement à celles traitées par l'un des trois phages (Gu et al. 2012). Chez l'homme, la plupart des études disponibles ont été faites il y a plusieurs décennies et n'ont pas

été conduites avec la rigueur méthodologique permettant de fournir la preuve d'efficacité clinique avec la rigueur actuelle (d'Herelle 1931; Tsulukidze 1941). Les publications d'essais cliniques récents sur la phagothérapie ne sont pas nombreuses pour plusieurs raisons. Les essais cliniques sont coûteux et les laboratoires pharmaceutiques ont pour l'instant peu d'intérêt à développer la phagothérapie, les essais sont donc souvent menés par des équipes académiques. En outre, les autorités de santé ont tardé à établir un cadre pour faciliter la mise en place d'études cliniques (Furfaro et al. 2018; Górski et al. 2018). Certaines études cliniques récentes ont toutefois été menées et ont révélé une efficacité plus ou moins importante de la phagothérapie. Par exemple, chez des patients atteints d'otites chroniques à *P.aeruginosa*, il a été montré l'efficacité de l'administration d'une dose de phages sur les symptômes de la maladie et sur le compte de bactéries (Wright et al. 2009). En revanche, l'utilisation d'un cocktail de phages par voie orale n'a pas permis d'améliorer l'état clinique d'enfants présentant une diarrhée nécessitant une hospitalisation, par rapport au traitement de référence (Sarker et al. 2016). Par ailleurs, l'étude récente PhagoBurn n'a pas permis de conclure à l'efficacité de la phagothérapie chez les grands brûlés (Jault et al. 2019). Les raisons mises en avant pour expliquer ces échecs thérapeutiques sont un mauvais choix de schéma thérapeutique ou de critères d'inclusion. En effet, la phagothérapie semble plus pertinente pour les infections lorsque le pathogène présente une multi-résistance aux antibiotiques, et une sensibilité au bactériophage administré. Cependant, la phagothérapie est utilisée pour traiter les infections bactériennes chez l'homme depuis plusieurs décennies en Europe de l'Est, notamment dans l'Institut Eliava, situé en Géorgie (Kutateladze and Adamia 2008).

Les bactériophages présentent un profil de toxicité intéressant en thérapeutique. Ils n'entraînent pas de lésion au niveau des tissus dans lesquels ils se distribuent. Ils ne sont pas capables d'infecter les cellules eucaryotes et ne présentent donc pas de reprotoxicité ou de mutagénicité sur les cellules humaines. Bien que les phages puissent induire une réaction immunitaire aiguë en cas de mauvaise purification de la préparation, celle-ci est généralement réduite ou inexistante, les phages faisant partie de notre environnement en permanence. Chez un modèle murin d'infection pulmonaire, il a été montré que la réponse immunitaire innée n'est pas plus importante par phagothérapie comparée à l'antibiothérapie, et cela pour deux couples dont 536/536_P1 (Dufour et al. 2019). Certaines études suggèrent la possibilité d'une réponse adaptative avec notamment production d'anticorps neutralisants dirigés contre les phages (Ochs et al. 1992; Rubinstein et al. 2000). Les principaux effets toxiques observés lors d'administration de phages sont dus à la libération de toxines lors de la lyse bactérienne. Il a

néanmoins été montré que la libération d'endotoxines n'était pas supérieure à celle induite par antibiothérapie (Matsuda et al. 2005). Cela a par exemple été montré pour le couple 536/536_P1 où l'exposition aux endotoxines était supérieure après traitement par céphalosporines ou par imipénème par rapport à la phagothérapie (Dufour et al. 2017). Toutes ces preuves de non toxicité des phages permettent un assouplissement progressif des conditions fixées par les autorités de santé pour les études cliniques des bactériophages et une prescription facilitée en usage compassionnel.

1.3.2. Modélisation de l'interaction phage/bactérie

Certains échecs des études de phagothérapie pourraient être causés par une mauvaise compréhension de la relation entre le schéma d'administration des phages et leur efficacité (Payne and Jansen 2001), conduisant à de mauvais choix de schémas d'administration des phages. Par exemple, si la dose de phages est trop faible ou si la voie d'administration conduit à une trop faible concentration de phages sur le lieu de l'infection, les bactéries peuvent continuer à se multiplier et entraîner des lésions avant que les phages soient en quantité suffisante pour contrôler la charge bactérienne. Le bactériophage étant un médicament, il est nécessaire de bien comprendre les relations entre dose, concentration et efficacité. Ceci peut être réalisé à l'aide de modèles mathématiques, plusieurs phénomènes sont alors à prendre en compte.

De la même manière qu'une molécule classique, la PK propre des phages (absorption, distribution et élimination) doit être étudiée. Cela se fait par administration d'une quantité de phages connue chez l'individu non infecté et suivi de la concentration de phages dans le sang et les tissus d'intérêt. Les bactériophages, contrairement aux molécules classiques, sont capables de se multiplier sur le lieu d'une infection. Les processus spécifiques de la multiplication virale sont donc à prendre en compte dans la manière de modéliser la dynamique des phages. La capacité des phages à s'adsorber aux bactéries, le temps nécessaire pour lyser une bactérie et le nombre de virus relargués lors de la lyse sont autant d'éléments à prendre en compte dans le modèle. Ils peuvent être estimés à partir de données d'expériences *in vitro* et *in vivo*, où phages et bactéries sont mis en contact. La capacité des bactériophages à se multiplier dépend de la probabilité de rencontre entre phages et bactéries et donc des concentrations bactérienne et virale sur le lieu d'infection. Il est donc nécessaire de prendre en compte la cinétique bactérienne. La multiplication « naturelle » des bactéries se quantifie idéalement en l'absence de traitement, en suivant l'évolution de la charge bactérienne sur le lieu d'infection.

En outre, pour certains couples phages/bactéries, le phénomène de résistance doit aussi être intégré au modèle (Skurnik and Strauch 2006).

Afin d'émettre des recommandations de dose, de voie et de nombre d'administrations de bactériophages à partir d'un modèle, il est donc crucial d'intégrer à la fois la cinétique bactérienne, la PK des phages et les interactions entre phages et bactéries. De plus, ces éléments font partie de la liste des informations qu'il est recommandé de fournir dans les études cliniques sur la phagothérapie (Abedon et al. 2011).

Bien que les premières données d'expérimentation de phages datent d'il y a plus de 100 ans, la littérature concernant la modélisation des phages est peu dense, en particulier pour la modélisation de l'interaction phage/bactérie. Pour comprendre ce type de modèles, il est nécessaire de détailler certains standards de modélisation de cinétique bactérienne d'une part, et de dynamique virale d'autre part. Ici, nous n'évoquerons pas la modélisation de la propagation des infections au niveau populationnel mais seulement la modélisation intra-hôte.

1.3.2.1. Modélisation de la cinétique bactérienne

Les modèles mathématiques de cinétique bactérienne, fréquemment écrits sous la forme d'équations différentielles, permettent de représenter et de quantifier l'effet des médicaments, le plus fréquemment des antibiotiques, sur la dynamique bactérienne ainsi que sur l'apparition de résistances.

Dans la littérature, la description de cinétiques bactériennes par modélisation est retrouvée plus fréquemment *in vitro* (Mouton et al. 1997; Nielsen et al. 2007) qu'*in vivo* (Guo et al. 2011). Ceci est explicable par plusieurs facteurs. Premièrement, lorsque les infections bactériennes sont localisées dans un tissu spécifique, il est souvent nécessaire de sacrifier l'animal pour relever la charge bactérienne sur le site d'infection. Dans ce contexte, la récolte de données longitudinales et donc la modélisation de l'évolution de la charge bactérienne est complexe. Deuxièmement, les interactions entre le système immunitaire de l'hôte et les bactéries sont aussi à prendre en compte, mais les marqueurs de l'immunité ne sont pas toujours suivis. Enfin, la variabilité inter-expérience et inter-individuelle est généralement beaucoup plus importante *in vivo* que dans les études *in vitro*. L'utilisation de modèles mixtes est intéressante parce qu'ils permettent de prendre en compte la variabilité, même lorsque les données sont éparpillées, comme expliqué section 1.1.1.

Beaucoup de modèles de cinétique bactérienne différents ont été publiés, une revue datant de 2013 liste les modèles les plus retrouvés en pharmacométrie des antibactériens (Nielsen and

Friberg 2013). Nous allons présenter les principaux modèles de cinétique bactérienne permettant de comprendre les travaux de ce doctorat.

Dans un système *in vitro* où les bactéries ont des ressources et de l'espace pour se multiplier, leur cinétique peut être décrite par l'équation 18 :

$$\frac{dB}{dt} = k_g B - k_d B, \quad (18)$$

où B est la charge bactérienne, k_g et k_d sont respectivement les paramètres de croissance et de mort naturelle bactérienne. Ici, la condition initiale $B(0)$ peut être équivalente à l'inoculum bactérien choisi I_B divisé par le volume de distribution estimé V_B (cette condition initiale peut aussi, plus simplement, être estimée par le paramètre B_0 , sans prendre en compte l'inoculum). Les paramètres k_g et k_d étant difficiles à identifier, l'écriture de l'équation de la cinétique bactérienne est alors fréquemment simplifiée (équation 19) :

$$\frac{dB}{dt} = \alpha B, \quad (19)$$

avec α , la croissance « nette » bactérienne équivalente à $k_g - k_d$.

Dans la plupart des expériences l'espace de croissance bactérienne est délimité, il est alors observé un plateau B_{max} , la cinétique bactérienne s'écrit alors (équation 20) :

$$\frac{dB}{dt} = \alpha B \left(1 - \frac{B}{B_{max}}\right). \quad (20)$$

Il existe différentes manières d'écrire la perturbation de cette cinétique par un antibactérien ABx (Figure 2, équation 21) :

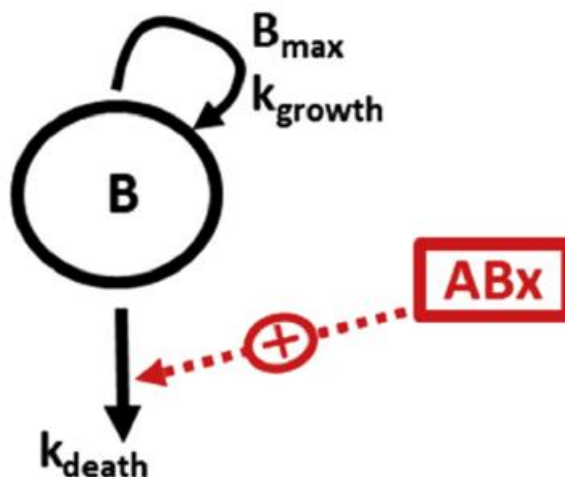


Figure 2. Représentation schématique d'un modèle de cinétique bactérienne (d'après Friberg and Guedj 2019).

$$\frac{dB}{dt} = \alpha B \left(1 - \frac{B}{B_{max}}\right) - ABx_{eff}B, \quad (21)$$

où ABx_{eff} est l'effet de cet antibactérien sur la cinétique bactérienne. Pour un modèle PKPD non-linéaire classique : $ABx_{eff} = \frac{E_{max}C}{EC_{50}+C}$ avec E_{max} l'effet maximal et EC_{50} la concentration C de l'antibactérien ABx qu'il est nécessaire d'atteindre pour obtenir la moitié de cet effet maximal. Dans certains cas, de manière moins mécanistique, cet effet peut être linéaire : $ABx_{eff} = \delta \times C$.

Lors de l'administration d'un médicament aussi, il est commun de pouvoir identifier l'apparition de souches résistantes, il est alors possible de décrire l'évolution de la charge bactérienne des deux sous-populations sensible (S , équation 22) et résistante (R , équation 23) :

$$\frac{dS}{dt} = \alpha S \left(1 - \frac{S}{B_{max}}\right) - ABx_{eff}S \quad (22)$$

$$\frac{dR}{dt} = w\alpha R \left(1 - \frac{R}{B_{max}}\right), \quad (23)$$

avec $B = R + S$, la charge bactérienne totale mesurée et w un paramètre de valeur sélective de populations résistantes par rapport aux sensibles (c'est-à-dire le ratio entre la vitesse nette de croissance des résistants par rapport à celle des sensibles). Les conditions initiales de ce système $S(0)$ et $R(0)$ peuvent être estimées (respectivement par les paramètres S_0 et R_0), elles peuvent aussi être paramétrées telles que $S(0) = \frac{I_B}{V_B} \times (1 - f)$ et $R(0) = \frac{I_B}{V_B} \times f$, avec f la fréquence de fluctuation, c'est-à-dire la proportion de bactéries résistantes en l'absence de traitement. Si les données à disposition le permettent, il est possible, à partir de modèles plus élaborés, de prendre en compte la pression de sélection due à l'antibactérien, c'est-à-dire le taux de mutation des souches sensibles afin qu'elles deviennent résistantes.

Lorsqu'on modélise la cinétique bactérienne *in vivo*, il peut être nécessaire de prendre en compte l'effet de l'immunité de l'hôte, en particulier l'immunité innée, médiée par les macrophages. Afin de prendre en compte cette composante dans le modèle, l'idéal est de suivre un ou des marqueurs de l'immunité de façon longitudinale, puis de considérer un système à plusieurs équations intégrant l'évolution de la charge bactérienne, celle de l'immunité et les interactions entre les deux (Roach et al. 2017). Lorsque aucun marqueur de l'immunité n'est

suivi durant l'étude, une façon simple de procéder est de considérer un temps nécessaire pour l'activation de l'immunité t_I (équations 24 et 25) :

$$\text{avant } t_I, \quad \frac{dB}{dt} = \alpha B \left(1 - \frac{B}{B_{max}}\right) \quad (24)$$

$$\text{après } t_I, \quad \frac{dB}{dt} = \alpha B \left(1 - \frac{B}{B_{max}}\right) - I_{eff}, \quad (25)$$

avec I_{eff} l'effet de l'immunité, qui peut s'écrire sous différentes formes.

Dans certaines études *in vivo*, si le décès est observé précocément pour un nombre conséquent d'individus, l'utilisation de modèles conjoints peut être un atout. En effet, la mortalité ici est liée à l'évolution de l'infection, les individus avec de hautes charges bactériennes ou un système immunitaire défaillant sont donc plus à risque de décès. Le risque est ici d'avoir une trop grande représentation des données d'individus survivants. Comme évoqué précédemment (voir section 1.1.1.2.), ceci peut biaiser l'estimation des paramètres d'un modèle incluant uniquement la description du marqueur longitudinal. L'utilisation d'un modèle conjoint (marqueur longitudinal-survie) permet de réduire ce biais.

1.3.2.2. Modélisation de la dynamique virale

Les bactériophages sont des virus dont les cellules cibles sont les bactéries. Afin de caractériser la dynamique des phages, l'utilisation des modèles classiques de dynamique virale, c'est-à-dire ceux développés pour les virus pathogènes, est nécessaire. Ces modèles ont initialement été développés pour des infections virales chroniques tels que le virus de l'immunodéficience humaine (Wei et al. 1995), puis ont été étendus à des infections aiguës telles que la grippe (Smith and Perelson 2011). Contrairement aux modèles de cinétique bactérienne, les modèles de dynamique virale prennent usuellement en compte la dynamique des cellules cibles des virus car les virus ont besoin de ces cellules pour se multiplier. La dynamique virale peut généralement être décrite par le système d'équations 26-28 (Perelson 2002) :

$$\frac{dS}{dt} = \lambda - dS - \beta VS \quad (26)$$

$$\frac{dI}{dt} = \beta VS - \delta I \quad (27)$$

$$\frac{dV}{dt} = pI - cV. \quad (28)$$

Ce modèle inclut à la fois les cellules cibles, aux stades non infectées (sensibles) S et infectées I , ainsi que la quantité de virus V . Les conditions initiales de ce système sont généralement $S(0) = \frac{\lambda}{d}$, $I(0) = 0$ et $V(0) = I_V$, avec I_V l'inoculum viral (si le virus est inoculé à $t = 0$). Dans ce système, les cellules cibles sont naturellement générées à la vitesse λ et subissent une élimination d'ordre 1 qui dépend du paramètre d . Le paramètre β est ici le taux d'infectivité, qui caractérise la vitesse d'infection et la probabilité de rencontre entre les virus et les cellules cibles. Les cellules infectées sont ensuite éliminées par le virus à la vitesse δ . De nouveaux virus sont produits à une vitesse p par les cellules infectées, puis sont éliminés naturellement à la vitesse c .

Le compartiment de cellules infectées est fréquemment séparé en deux compartiments I_1 et I_2 , avec les cellules I_1 qui sont latentes (non productrices) et I_2 productrices de virus (Figure 3, équations 29-32).

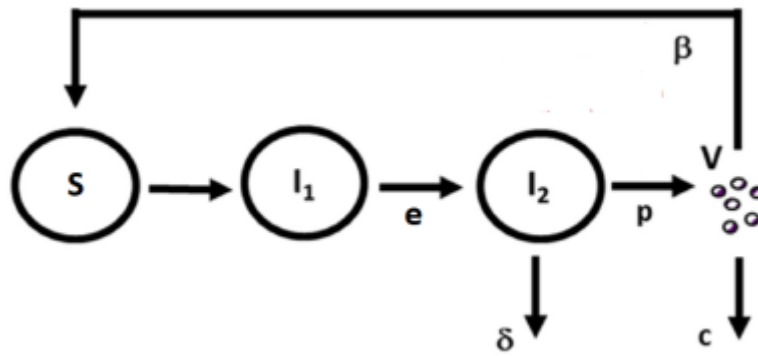


Figure 3. Représentation schématique d'un modèle de dynamique virale (d'après Friberg and Guedj 2019).

$$\frac{dS}{dt} = \lambda - dS - \beta VS \quad (29)$$

$$\frac{dI_1}{dt} = \beta VS - eI_1 \quad (30)$$

$$\frac{dI_2}{dt} = eI_1 - \delta I_2 \quad (31)$$

$$\frac{dV}{dt} = pI_2 - cV. \quad (32)$$

Ce modèle permet de prendre en compte le délai nécessaire au virus pour « utiliser » la cellule cible avant que celle-ci puisse produire de nouveaux virions. Le paramètre e caractérise ce

délai, mais l'estimation de ce paramètre est souvent difficile. Des expériences *in vitro* peuvent néanmoins aider à le déterminer.

D'autres modèles de dynamique virale plus élaborés ont été développés, incluant notamment des effets de traitement antiviraux, voire de cellules réfractaires. Enfin, comme en cinétique bactérienne, intégrer des marqueurs de la réponse immunitaire peut aider à expliquer les phénomènes observés *in vivo* (Friberg and Guedj 2019). Par exemple, dans le cadre d'une étude sur la maladie à virus Ebola chez le macaque, les lymphocytes T et trois cytokines ont été mesurés de manière longitudinale et intégrés au système d'équations. De plus, le lien entre charge virale, marqueurs immunologiques et survie a été montré par l'utilisation d'un modèle conjoint (Madelain et al. 2018).

1.3.2.3. Modèles mathématiques existants pour la description de l'interaction phage/bactérie

Les premiers modèles mathématiques des bactériophages reprenant les standards de cinétique bactérienne et de dynamique virale ont été développés à partir des années 2000 (Payne 2000). A partir d'un modèle composé d'un système de trois équations différentielles : bactéries non infectées S (équation 33), bactéries infectées I (équation 34) et phages V (équation 35) :

$$\frac{dS}{dt} = \alpha S - \beta VS \quad (33)$$

$$\frac{dI}{dt} = \alpha I + \beta VS - \delta I \quad (34)$$

$$\frac{dV}{dt} = b\delta I - \beta VS - k_e V. \quad (35)$$

Dans ce modèle, les conditions initiales sont généralement $S(0) = S_0$, $I(0) = 0$ et $V(0) = 0$, lorsque les phages sont administrés après le début du suivi. Ce modèle a permis d'avancer l'importance du temps d'inoculation des phages pour le contrôle de l'infection bactérienne. Les paramètres clés caractérisant les interactions entre phages et bactéries sont le taux d'infectivité β et la vitesse de lyse δ . Le paramètre b , généralement absent en dynamique virale classique, caractérise le *burst size*, c'est-à-dire le nombre de virions libérés à la fin du cycle d'infection, lors de la lyse de la bactérie. Ce premier modèle mathématique constitue une base théorique pour la modélisation de l'interaction phage-bactérie. Il est bien adapté à la description d'expériences *in vitro* (Cairns et al. 2009). Il n'inclut cependant pas certaines complexités que l'on peut rencontrer lors de l'utilisation de phages *in vivo* (Weld et al. 2004). Quelques années

plus tard, un modèle étendait ces équations pour intégrer l'émergence de souches bactériennes résistantes au phage et une expansion bactérienne limitée par les ressources disponibles dans le système (Levin and Bull 2004). Certains modèles intègrent aussi la pression de sélection exercée par les bactériophages, par le biais d'un paramètre de taux de mutation des bactéries sensibles vers des bactéries résistantes (Cairns et al. 2009). Un modèle récent intègre aussi la dynamique de l'immunité de l'hôte et son effet sur la charge bactérienne (Roach et al. 2017). La prochaine étape pourrait être l'utilisation de modèles mixtes, de manière à intégrer et expliquer certaines variabilités observées *in vivo*. L'utilisation de modèles conjoints, c'est-à-dire liant la survie à la cinétique bactérienne ou aux marqueurs de l'immunité est aussi une piste à explorer. La modélisation permet une meilleure compréhension et la quantification de l'interaction phage-bactérie, et pourrait *in fine* aider au choix du schéma d'administration de la phagothérapie chez l'homme. Un objectif de ces travaux de thèse est d'utiliser les modèles mixtes afin de caractériser l'interaction entre bactériophages et bactéries (section 3.).

1.4. Objectifs de la thèse

La première partie de ce doctorat porte sur l'évaluation et l'optimisation de protocoles des études longitudinales analysées par modèles non linéaires à effets mixtes. Parmi les travaux de recherche réalisés, deux sont des développements de méthodes pour la planification de protocoles basée sur l'utilisation de la matrice d'information de Fisher et un autre travail porte sur la comparaison de protocoles par simulations, appliquée en oncologie.

Le premier travail s'inscrit dans la continuité de travaux de recherche menés au sein de notre équipe ; d'une part dans le développement de méthodes d'évaluation de l'information de Fisher adaptées aux modèles de données non continues (section 1.2.1.1.), et d'autre part dans le développement de méthodes de protocoles robustes, capables de prendre en compte l'incertitude du modèle et/ou des valeurs de paramètres (section 1.2.1.3.). Dans le cadre de l'analyse de données binaires répétées, notre objectif était donc de développer une méthode de planification prenant en compte l'incertitude du modèle, en évaluant la matrice d'information de Fisher par méthode de Monte Carlo-Monte Carlo Hamiltonien. Un objectif secondaire était d'évaluer la pertinence de l'approche développée, c'est pourquoi nous avons réalisé des simulations d'essais cliniques (section 2.1).

Nous avons vu précédemment la nécessité de développer de nouveaux algorithmes à l'optimisation de protocoles pour les modèles mixtes à partir de la matrice d'information de Fisher (section 1.2.1.4). C'est pourquoi l'objectif du deuxième travail de recherche était

d'adapter les algorithmes multiplicatifs (jusqu'ici appliqués aux régressions linéaires ou non-linéaires simples) à l'optimisation de protocole pour modèles mixtes. Un autre objectif était de l'inclure dans la prochaine version de PFIM (section 1.2.1.4). Ces travaux sont présentés dans la section 2.2.

Nous avons vu en section 1.1.5. que la modélisation de la taille tumorale mesurée de manière longitudinale était de plus en plus développée, en particulier dans le développement d'anticancéreux utilisés en combinaison. Nous avons d'autre part vu que les simulations d'essais cliniques étaient adaptées à la comparaison de différents protocoles d'études (section 1.2.2.). Le troisième travail de ce doctorat porte sur la comparaison de protocoles en phase précoce de développement clinique en oncologie. Pour cette application, l'objectif était de savoir quels étaient les avantages et inconvénients de différents protocoles dans le cadre d'une étude de combinaison d'anticancéreux, avec comme critère de décision la décroissance tumorale précoce. Par simulation d'essais cliniques, des protocoles à un bras (comparaison à une référence historique), deux bras (essai clinique randomisé classique) et quatre bras (avec plusieurs doses possibles pour un médicament) ont donc été comparés, principalement en termes de puissance statistique (et d'erreur de type I) pour détecter un effet de la combinaison par rapport au traitement de référence. Ce travail est présenté dans la section 2.3.

L'autre partie de ce doctorat porte sur la phagothérapie, pour laquelle nous avons vu que la modélisation pouvait permettre de mieux comprendre l'interaction entre bactériophages et bactéries (section 1.3.2.). Des études ont été conduites *in vitro* et *in vivo* sur la souche bactérienne *Escherichia coli* 536 et le phage 536_P1 afin de comprendre les interactions entre bactéries et phages, ainsi que le lien entre le schéma d'administration des phages et leur efficacité. Notre travail de recherche avait pour objectif de construire un modèle mathématique permettant de caractériser l'évolution de la concentration de phage et de la charge bactérienne et d'en estimer les paramètres. Ce travail est présenté dans la troisième partie du manuscrit.

Une discussion de l'ensemble des travaux de cette thèse est proposée en quatrième partie.

Deuxième partie : Evaluation et optimisation de protocoles dans les modèles mixtes

Dans cette partie, nous présentons les travaux autour de l'évaluation et l'optimisation de protocoles d'études longitudinales analysées par modèles mixtes. En section 2.1. est abordé le développement d'une méthode de planification de protocole robuste prenant en compte l'incertitude dans le choix du modèle pour des données binaires répétées. La section 2.2. traite de l'extension d'un algorithme multiplicatif à l'optimisation de protocoles pour les modèles mixtes, ainsi que de son intégration à la version de PFIM actuellement en développement. Enfin, en section 2.3., nous proposons une comparaison de protocoles par simulations appliquée à l'étude d'une combinaison d'anticancéreux.

2.1. Protocoles robustes tenant compte de l'incertitude du modèle pour réponse binaire répétée

2.1.1. Résumé

Nous avons proposé une méthode pour la planification de protocole dans le cadre d'études avec réponse binaire répétée et analysée par modèles mixtes. Cette approche est basée sur la FIM évaluée par échantillonnage de Monte Carlo et de Monte Carlo Hamiltonien, méthode existante dans la librairie R *MIXFIM* (Riviere and Mentré 2015; Riviere et al. 2016). En planification de protocoles, il est nécessaire de spécifier le modèle *a priori*. Or, si celui-ci est très différent du modèle finalement utilisé pour décrire les données, le protocole choisi pourra être sous-optimal. Nous avons donc proposé une méthode robuste tenant compte de l'incertitude du modèle en s'appuyant sur plusieurs modèles candidats.

Nous avons appliqué cette approche dans le cadre d'une étude avec deux groupes d'individus (placebo & traité) et supposé des modèles logistiques mixtes. Chaque protocole était composé de quatre temps de mesures, dont un au début de l'étude et un à 12 mois, c'est-à-dire la fin de l'étude. Nous avons cherché à optimiser, de façon combinatoire, l'allocation des deux autres temps de mesures parmi les valeurs possibles suivantes : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 et 11 mois. Premièrement, nous avons réalisé une planification de protocole locale (non robuste) pour chacun des 4 modèles candidats et selon 3 critères d'optimalité différents. Le critère de D-optimalité, visant à assurer une précision globale de l'ensemble des paramètres, a été utilisé. Le

critère D_S a aussi été employé dans le but d'optimiser la précision d'estimation d'un sous-ensemble de paramètres, dans notre cas, l'effet de la covariable « traitement ». Cet effet traitement étant détecté par test de Wald, l'optimisation de sa précision d'estimation permet d'augmenter la puissance de ce test. Enfin, nous avons proposé l'utilisation du critère DD_S qui permet d'assurer un compromis entre la précision d'estimation de l'ensemble des paramètres et la puissance du test de Wald pour détecter l'effet traitement. Deuxièmement, afin de prendre en compte l'incertitude du modèle, nous avons alloué un poids à chacun des modèles candidats compris entre 0 et 1, poids qui représente l'aptitude que l'on donne à un modèle à être celui qui sera utilisé pour décrire les données. Dans notre exemple, nous avons considéré le même poids (0,25) pour chacun des quatre modèles candidats. Nous avons utilisé ces poids dans le calcul des différents critères robustes ou composés (CD pour *Compound D*, CD_S et CDD_S) afin d'optimiser les temps de mesures. Nous avons aussi utilisé la FIM afin de prédire la puissance statistique moyenne avec ces modèles candidats, pour détecter l'effet traitement.

Les temps de mesures obtenus selon les modèles et selon les critères sont différents. Les D -, D_S - et DD_S -efficacités ont été calculées avec les différents protocoles. L'approche robuste permet d'éviter des protocoles pouvant être peu efficaces sous certains modèles. Par exemple, le protocole CD-optimal conduit à une D -efficacité supérieure à 0.8 avec n'importe lequel des modèles candidats mais la planification locale, couplée à un mauvais choix de modèle *a priori*, peut conduire à une D -efficacité inférieure à 0.65. Nous avons prédit le nombre de sujets nécessaires (NSN) pour obtenir une puissance de 90 % permettant de détecter l'effet du traitement avec le protocole robuste (CDD_S -optimal) $\xi_{CDD_S} = (0,4,11,12)$, un protocole planifié avec une approche non robuste (DD_S -optimal pour le modèle le plus simple) $\xi_{DD_S,1} = (0,2,11,12)$ et un protocole non optimisé (avec des temps de mesures équi-répartis) $\xi_{ES} = (0,4,8,12)$: ils étaient respectivement de 274, 320 et 358.

Cinq cents jeux de données ont été simulés sous chaque modèle candidat et avec les trois protocoles différents : robuste, DD_S -optimal, et non optimisé. Les paramètres ont ensuite été estimés pour chaque jeu de données et le bon modèle (correspondant au modèle de simulation), avec l'algorithme SAEM de Monolix 2016R1. Les résultats obtenus avec les trois protocoles ont été comparés en termes de biais, de précision d'estimation des paramètres, d'erreur de type I et de puissance. Les simulations ont confirmé que l'utilisation d'un protocole robuste permet une précision d'estimation acceptable des paramètres ainsi qu'une puissance de tests supérieure à 80 % avec chaque modèle.

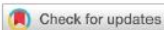
Nous avons aussi appliqué ces méthodes à une étude réelle de données binaires répétées provenant de 4162 sujets de la cohorte EPESE (*Established Populations for Epidemiologic Studies of the Elderly* (Landerman et al. 2011)) et analysées par modèle logistique mixte. Nous avons montré que les temps de mesures et la taille de l'étude auraient pu être choisis de manière plus efficace afin de détecter des effets de covariables, par exemple en optimisant les temps de mesures selon le critère DD_5 .

La stratégie proposée dans ce travail est pertinente pour la planification de protocole d'études longitudinales en cas d'incertitude du modèle. Ce travail a fait l'objet d'un article publié dans *Statistical Methods in Medical Research*.

2.1.2. Article 1 (publié)

Robust designs accounting for model uncertainty in longitudinal studies for binary outcome, publié dans *Statistical Methods in Medical Research*, 2020

Jérémy Seurat, Thu Thuy Nguyen, France Mentré



Article

Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes

Jérémy Seurat,  Thu Thuy Nguyen and France MentréStatistical Methods in Medical Research
2020, Vol. 29(3) 934–952

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0962280219850588

journals.sagepub.com/home/smm



Abstract

To optimize designs for longitudinal studies analyzed by mixed-effect models with binary outcomes, the Fisher information matrix can be used. Optimal design approaches, however, require a priori knowledge of the model. We aim to propose, for the first time, a robust design approach accounting for model uncertainty in longitudinal trials with two treatment groups, assuming mixed-effect logistic models. To optimize designs given one model, we compute several optimality criteria based on Fisher information matrix evaluated by the new approach based on Monte-Carlo/Hamiltonian Monte-Carlo. We propose to use the DD_5 -optimality criterion, as it ensures a compromise between the precision of estimation of the parameters, and hence the Wald test power, and the overall precision of parameter estimation. To account for model uncertainty, we assume candidate models with their respective weights. We compute robust design across these models using compound DD_5 -optimality. Using the Fisher information matrix, we propose to predict the average power over these models. Evaluating this approach by clinical trial simulations, we show that the robust design is efficient across all models, allowing one to achieve good power of test. The proposed design strategy is a new and relevant approach to design longitudinal studies with binary outcomes, accounting for model uncertainty.

Keywords

Optimal design, compound optimality, Fisher information matrix, longitudinal binary data, nonlinear mixed effect models

1 Introduction

Nonlinear mixed effect models (NLMEMs) are frequently used in model-based drug development to analyze pharmacokinetic/pharmacodynamic data.¹ The use of NLMEMs (i.e. the population approach) allows the estimation of mean parameters, inter and intra-subject variabilities as well as covariate effects and is appropriate for exploiting the richness of longitudinal data.^{2–4} NLMEMs can compensate for the lack of individual information by borrowing the strength from the data in the whole population and therefore allow for precise parameter estimation even with sparse designs, where few samples are collected from each subject.

NLMEMs are increasingly used for the analysis of longitudinal clinical studies, for both continuous and discrete data (binary, count or time-to-event). Binary outcomes are frequently encountered to characterize the clinical response in different therapeutic areas such as infectiology (virological success, bacterial carrying), and can be used for responder/non-responder analysis as well as for toxicity analysis. For instance, the carriage of *Streptococcus pneumoniae* was studied in children between two months and two years, by collecting nasopharyngeal swabs and aspirates over 10 visits to detect the impact of determinants as age or health status on the proportion of positive bacterial samples.⁵ In another context, Mestre et al.⁶ recently proposed a binary outcome (corresponding to a minimal change in UDysRS Part III Impairment scale from baseline) to characterize the remission of dyskinesia in Parkinson's disease. The analysis of the longitudinal data in these kinds of outcomes required discrete response generalized linear mixed effect models (GLMMs) or discrete response NLMEMs and adequate estimation methods.

IAME, UMR 1137, INSERM, Université Paris Diderot, Paris, France

Corresponding author:

Jérémy Seurat, IAME, UMR 1137, INSERM, University Paris Diderot Sorbonne Paris Cité, 16, rue Henri Huchard, 75018 Paris, France.

Email: jeremy.seurat@inserm.fr

Before modeling, it is crucial to choose an appropriate design in order to obtain good precision of parameter estimates. Indeed, the informativeness of a dataset for parameter estimation depends on the design choice. A design in NLMEMs, called a population design, is composed of the number of elementary designs, the specification of each elementary design, and the associated number of subjects. In these settings, the term elementary design is used to describe a group of subjects with identical design characteristics such as the number of samples per subject and the allocation of informative times and doses. To evaluate and optimize designs, two approaches have been proposed. The first approach, based on clinical trial simulation (CTS), is very time-consuming and is therefore limited in term of designs that can be evaluated. Alternatively, the expected Fisher information matrix (FIM) can be used,⁷⁻⁹ as its inverse is defined as the lower bound of the variance-covariance matrix of any unbiased estimated parameters, according to the Cramér-Rao inequality. However, the FIM has no analytical form in NLMEMs and its computation, which requires multiple integrations, can be challenging. A first method to evaluate the FIM in NLMEMs, based on first order (FO) linearization of the model around the expectation of random effects, was proposed¹⁰ and implemented in several software programs.¹¹ Although efficient in general, FO presents limitations when used with complex models, large variability^{12,13} and when using longitudinal models for discrete data.¹³ Alternative to compute the FIM in continuous NLMEMs, without linearization, has been proposed using Monte Carlo (MC) integrations.¹⁴ There are also a large body of work on the FIM for GLMMs with random intercepts, using marginal quasi-likelihood, penalized quasi-likelihood, complete enumeration-based or MC methods.^{15,16} More recently, new methods have been developed to compute the FIM for both continuous and discrete NLMEMs using MC methods combined with adaptive Gaussian quadrature (AGQ)¹⁷ or Hamiltonian Monte Carlo (HMC).¹³ The latter approach (MC/HMC) was implemented in the R package *MIXFIM*,¹⁸ which uses functions written in the probabilistic language Stan,¹⁹ which was developed for Bayesian inference. This method, by efficiently drawing HMC samples and calculating partial derivatives of the log-likelihood, has been shown to be more suitable than MC/AGQ in complex models with a large number of random effects.¹³

From the expression of the FIM evaluated by these approaches, different optimality criteria can then be computed to optimize designs. The widely used D-optimality criterion consists in maximizing the determinant of the FIM, i.e. to optimize the precision of estimation for all model parameters. As an alternative to the D-optimality, the D_S -optimality²⁰ can accommodate situations in which only a subset of the model parameters is of interest. This can be particularly useful to optimize the precision of a “treatment effect” parameter on the longitudinal evolution of biomarkers in clinical trials for example, which is directly linked to the power of the Wald test to detect this effect.²¹ However, D_S -optimal designs can lead to problem in estimation of all the parameters of the model because of lack of experimental identifiability.²² To find a balance between optimizing the precision of the parameters to be tested and the precision all the model parameters, a mixture of D- and D_S -optimality, the DD_S -optimality, can be used.²⁰ However, these methods require a priori knowledge of the model and its parameter values, which can usually be obtained from previous studies but lead to designs that are locally optimal. If the final model is very different from the a priori assumed model, the design optimized with the wrongly assumed model might not be informative enough to precisely estimate parameters of the final model. Therefore, there is need to develop and evaluate new methods to optimize robust designs accounting for model uncertainty. In order to propose optimal designs across a set of candidate models, we make use of both the theory of compound optimal design⁷ and the principle of model averaging.²³ The compound optimal design theory was previously used to propose informative designs for both estimation and model discrimination,²⁴ and recently to find a common design for cocktail of drugs.²⁵ Meanwhile, model averaging is a promising alternative approach to model selection when modeling data, which uses model weights (computed from AIC or BIC) to calculate a weighted average of the predictions.²⁶⁻²⁸

Combining these approaches, our main objective is to propose and evaluate a new methodological strategy for robust designs in longitudinal studies with discrete outcomes taking into account model uncertainty, based on the expected FIM in NLMEMs. We choose to use the MC/HMC method to compute the FIM without any linearization. From expression of the FIM, we predict the power of the Wald test to detect a treatment effect included in the model and to calculate the number of subjects needed for a required power. Our FIM-based strategy aims to find designs that are both informative to achieve good average power over several models and to precisely estimate the whole set of population parameters. We illustrate these methods through an example of designing a trial with repeated binary outcomes, inspired from a previous work within the team,²⁹ considering several candidate models which include a treatment effect on the longitudinal response. The relevance of these approaches is evaluated by CTS in terms of bias and imprecision of parameter estimates and power of the Wald test.

In Section 2, we detail the notations and present different optimality criteria for a given model or for taking into account model uncertainty. In Section 3, we show an application of these methods to design a longitudinal trial with binary outcomes. In Section 4, we present a simulation study to evaluate the relevance of the proposed design strategy. In Section 5, we show an application to a real study. Finally, we discuss the results and perspectives of this work in Section 6.

2 Methods

2.1 Basic concepts and notation

2.1.1 Population design

The elementary design ξ_i for the individual i ($i = 1, \dots, N$) is defined by the number n_i of samples and their allocation in time $(t_{i1}, \dots, t_{in_i})$. The population design $\Xi = \{N, (\xi_1, \dots, \xi_N)\}$ is defined by the number of individuals N and the set of elementary designs to be performed in each individual (ξ_1, \dots, ξ_N) , with a total number of observations $\sum_{i=1}^N n_i$.

2.1.2 NLMEMs for binary outcomes

This work considers a set of candidate NLMEMs $m = 1, \dots, M$ for binary data. Considering a given logistic model m , the logit of the conditional probability for observation y_{ij} of individual i at sample j ($j = 1, \dots, n_i$) is written

$$\text{logit}(p(y_{ij} = 1|b_i)) = f_m(t_{ij}, g(\mu_m, b_i, z_i, \beta_m)) \quad (1)$$

In equation (1), f_m is the function describing the logit-probability of y_{ij} given individual random effects b_i , for a time t_{ij} and a vector of subject-specific parameters modelled through g . The function g depends on the vector of fixed effects μ_m, b_i as well as the vector of covariates z_i and the vector of covariate effects β_m . The random effects are assumed to follow a normal distribution with mean zero and variance-covariance matrix Ω_m which accounts for the inter-individual variability, i.e. $b_i \sim N(0, \Omega_m)$. We denote by ψ_m the vector of population parameters, i.e. $\psi_m = (\mu_m^T, \beta_m^T, \Omega_{m,u}^T)^T$, where $\Omega_{m,u}$ is a vector containing all unique elements of Ω_m . Let P_m denote the length of the vector ψ_m . Observations are usually assumed to be independent conditionally upon the random effects, i.e. the joint conditional probability for the vector of observations for individual i ($y_i = (y_{i1}^T, \dots, y_{in_i}^T)^T$) is

$$p(y_i|b_i, \psi_m) = \prod_{j=1}^{n_i} \frac{\exp(f_m(t_{ij}, g(\mu_m, b_i, z_i, \beta_m)))}{1 + \exp(f_m(t_{ij}, g(\mu_m, b_i, z_i, \beta_m)))}$$

2.1.3 FIM

For a model m and its parameter values ψ_m , if the N subjects are independent, the population FIM, $\mathcal{M}(\psi_m, \Xi)$, is the sum of N elementary FIMs, $\mathcal{M}(\psi_m, \xi_i)$, i.e. $\mathcal{M}(\psi_m, \Xi) = \sum_{i=1}^N \mathcal{M}(\psi_m, \xi_i)$. In this work, we assume the same elementary design for all individuals, i.e. $\xi_i = \xi$ for $i = 1, \dots, N$, then $\Xi = \{N, \xi\}$ and $\mathcal{M}(\psi_m, \Xi) = N\mathcal{M}(\psi_m, \xi)$. The elementary FIM for ψ_m can be expressed as

$$\mathcal{M}(\psi_m, \xi) = E_y \left(\frac{\partial \log(L(y|\psi_m))}{\partial \psi_m} \frac{\partial \log(L(y|\psi_m))^T}{\partial \psi_m} \right) \quad (2)$$

where the likelihood L of the observations vector y of an individual (subscript i is omitted for simplicity) is given by

$$L(y|\psi_m) = \int_b p(y|b, \psi_m) p(b|\psi_m) db \quad (3)$$

with $p(y|b, \psi_m)$ the probability density function (p.d.f.) of y given the random effects b , and $p(b|\psi_m)$ the p.d.f. of b . To evaluate the FIM in equation (2), one needs to compute two integrations: one over the observations y and one in equation (3) over the distribution of random effects. We evaluate the former by MC integrations and the latter by HMC, as proposed in Riviere et al.,¹³ using the package *MIXFIM*¹⁸ in R v3.2.1. *MIXFIM* calls functions of the R package *rstan* written in the probabilistic language Stan¹⁹ to draw HMC samples and calculate partial derivatives of the log-likelihood.

2.2 Optimality criteria

The methods we propose to find robust designs are presented in this section. Different criteria to find an optimal design are presented, according to different purposes. For a given model, based on the determinant of the FIM, the D-optimality is used to obtain informative designs for a precise estimation of all parameters of a model, the D_S-optimality accommodates situations in which only a subset of parameters is of interest, and the compound DD_S-optimality offers a compromise between the D- and the D_S-criteria. Averaging over several models and weighting each one, the compound CD-, CD_S- and CDD_S-optimalities accounts for model uncertainty.

2.2.1 Optimality criteria for a given model

To optimize the estimation precision for the whole set of population parameters ψ_m , we maximize the widely used⁷ D-optimality criterion $\Phi_{D,m}(\Xi)$ (equation (4), Table 1).

Let $\psi_{S,m}$, a subset of ψ_m , be the vector of parameters of interest of length S_m . To optimize the estimation precision for the S_m parameters, we use the D_S-optimality $\Phi_{D_S,m}(\Xi)$ (equation (5), Table 1).⁷ This criterion is particularly useful to find a design which maximizes the predicted power of the Wald test or minimizes the number of subjects needed to detect a significant treatment effect β_m , by focusing on the precision of estimation for β_m .

Let $\psi_{T,m}$ be the remaining parameters of ψ_m beside $\psi_{S,m}$. To find a compromise between precision of $\psi_{S,m}$ and $\psi_{T,m}$, the compound DD_S-optimality criterion,^{7,20} $\Phi_{DD_S,m}(\Xi, \alpha_m)$ is used (equation (6), Table 1). α_m is a term between 0 and 1, which quantifies the interest in the estimation precision of $\psi_{S,m}$. When $\alpha_m = 1/P_m$, the DD_S-criterion coincides with the D-criterion; thus, all the parameters of the model are of equal interest. When $\alpha_m = 1$, the compound DD_S-criterion coincides with the D_S-criterion. When $\alpha_m = 0$, the precision of the parameters $\psi_{T,m}$ is optimized. Maximizing the DD_S-criterion with $1/P_m < \alpha_m < 1$ provides a compromise between the D- and the D_S-criteria.

For any given optimality criterion X , the relative efficiency $E_{X,m}(\Xi)$ of a design Ξ , with respect to the X -optimal design $\Xi_{X,m}$ for model m , is computed as

$$E_{X,m}(\Xi) = \frac{\Phi_{X,m}(\Xi)}{\Phi_{X,m}(\Xi_{X,m})}$$

2.2.2 Robust optimality criteria accounting for model uncertainty

We propose to account for uncertainty in the model choice by computing optimality criteria based on the theory of compound optimality⁷ and of model averaging.²³ We assume that a set of M candidate models is available, with each model associated to a weight w_m ($\sum_{m=1}^M w_m = 1$) quantifying the balance between the M models. To find a common optimal design for the M models, according to the X -optimality, we maximize the weighted product of

Table 1. Various optimality criteria used depending on parameters of interest and accounting or not for model uncertainty.

Parameters of interest	Given model m	Set of candidate models $m = 1, \dots, M$
All the parameters ψ_m	D-optimality $\Phi_{D,m}(\Xi) = \text{Det}(\mathcal{M}(\psi_m, \Xi))^{1/P_m}$ (4)	CD-optimality ^{7,24,25} $\Phi_{CD}(\Xi) = \prod_{m=1}^M (\Phi_{D,m}(\Xi))^{w_m}$ (7)
Subset of parameters $\psi_{S,m}$	D _S -optimality $\Phi_{D_S,m}(\Xi) = \left(\frac{\text{Det}(\mathcal{M}(\psi_m, \Xi))}{\text{Det}(\mathcal{M}(\psi_{T,m}, \Xi))} \right)^{1/S_m}$ (5)	CD _S -optimality $\Phi_{CD_S}(\Xi) = \prod_{m=1}^M (\Phi_{D_S,m}(\Xi))^{w_m}$ (8)
Compromise between $\psi_{S,m}$ and $\psi_{T,m}$	DD _S -optimality $\Phi_{DD_S,m}(\Xi, \alpha_m) = \left(\text{Det}(\mathcal{M}(\psi_{T,m}, \Xi)) \right)^{\frac{1-\alpha_m}{P_m-S_m}} \times \left(\frac{\text{Det}(\mathcal{M}(\psi_m, \Xi))}{\text{Det}(\mathcal{M}(\psi_{T,m}, \Xi))} \right)^{\frac{\alpha_m}{S_m}}$ (6)	CDD _S -optimality $\Phi_{CDD_S}(\Xi) = \prod_{m=1}^M (\Phi_{DD_S,m}(\Xi))^{w_m}$ (9)

Note: ψ_m is the vector of population parameters (of size P_m) for a model m ; $\psi_{S,m}$ the subset of parameters of interest (of size S_m) and $\psi_{T,m}$ which are not of interest. $\mathcal{M}(\psi_m, \Xi)$ is the matrix containing the Fisher information for all the parameters ψ_m and $\mathcal{M}(\psi_{T,m}, \Xi)$ is obtained by truncation of $\mathcal{M}(\psi_m, \Xi)$ and keeping only the rows and columns corresponding to $\psi_{T,m}$. α_m ($0 \leq \alpha_m \leq 1$) is a term which quantifies the balance between the D- and the D_S-optimality criteria and expresses the interest in the precision of estimation for $\psi_{S,m}$. w_m is the weight for the model m ($\sum_{m=1}^M w_m = 1$).

efficiencies $\prod_{m=1}^M E_{X,m} w_m$, with X being the D-, D_S - or DD_S -optimality criterion. Maximizing the weighted product of efficiencies is equivalent to maximizing the weighted product of the criteria, hence the definition of compound optimality criteria,⁷ $\Phi_{CD}(\Xi)$, $\Phi_{CD_S}(\Xi)$ or $\Phi_{CDD_S}(\Xi, \alpha_m)$ as provided (equations (7) to (9), Table 1). Of note, it is also equivalent to minimizing the weighted sum of minus-log-criteria, that is a convex function to which the General Equivalence Theorem can be applied.³⁰

3 Application to design a longitudinal trial with binary response

In this section, we present how the methods introduced in Section 2 are applied to design a longitudinal study over one year, with two balanced treatment groups and binary response. Candidate NLMEMs including a treatment effect and describing the response probability over time are considered, inspired by Lestini et al.²⁹ and Ogungbenro and Aarons.³¹ We aim to determine the optimal location of a limited number of measurement times chosen among a discrete design composed of (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12) in month unit, according to different optimality criteria. We investigate the predicted efficiency of the various optimal designs obtained as well as the power of the Wald test to detect a treatment effect over all the candidate models.

3.1 Models and parameters

The studied models describe binary responses ($y_{ij} = 1$, responder or 0, non responder) recorded over time ($t_{ij} = 0$ to 12 months) in two different treatment groups. The logit of response probability at time t_{ij} is given by equation (1).

We denote $f_m(t_{ij}) = f_m(t_{ij}, g(\mu_m, b_i, z_i, \beta_m))$ and omit index ij for simplicity in the followings. We consider four candidate models f_m describing the evolution of the logit-probability of the response over time (Figure 1), which are the linear model (M1), loglinear model (M2), quadratic model (M3) and exponential model (M4).

$$\begin{aligned} \text{M1 : } f_1(t) &= \theta_1 + \theta_2(1 + \beta \times 1_T)t \\ \text{M2 : } f_2(t) &= \theta_1 + \theta_2(1 + \beta \times 1_T)\log(t + 1) \\ \text{M3 : } f_3(t) &= \theta_1 + \theta_2(1 + \beta \times 1_T)t^2 \\ \text{M4 : } f_4(t) &= \theta_1 + \theta_2(1 + \beta \times 1_T)(\exp(\theta_3 t) - 1) \end{aligned}$$

1_T is a treatment group indicator variable (with $1_T = 0$ if control group, $1_T = 1$ if treated group). We denote β the effect size of the treatment on parameter θ_2 . Model M1 and its parameters are inspired by Lestini et al.²⁹ Values of ψ_1 are given in Table 2, θ_1 and θ_2 follow a normal distribution:

$\theta_p = \mu_p + b_p$ where $b_p \sim N(0, \omega_p^2)$ for $p = \{1, 2\}$. Models M2 to M4 are alternative models with parameters normally distributed as in M1. Values of their fixed parameters are chosen to give the same mean value of the logit of the response probability as for M1 at the beginning ($t = 0$ month) and the end ($t = 12$ months) of the study in the two treatment groups. Similar treatment effect value and a standard deviation ω_2 equal to 188% of μ_2 for are assumed for all models. Assumed values of population parameters ψ_m for each model m ($m = 1, \dots, 4$) are given in Table 2.

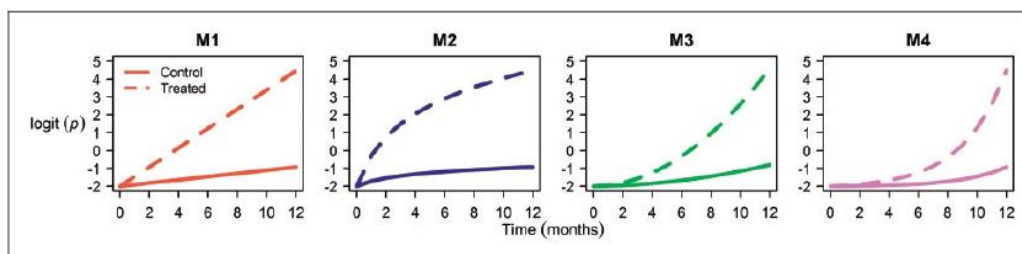


Figure 1. Plots of the logit of the response probability (p) over 12 months in each treatment group for the four candidate models: M1 linear (in orange), M2 loglinear (in blue), M3 quadratic (in green) and M4 exponential (in purple), using the fixed effect values indicated in Table 2. Solid lines represent the control group and dashed lines the treated group.

Table 2. Population parameter values ψ_m for each candidate model m ($m = 1, \dots, 4$): M1 is the linear model, M2 the loglinear model, M3 the quadratic model and M4 the exponential model.

	M1	M2	M3	M4
μ_1	-2	-2	-2	-2
μ_2	0.09	0.42	7.50×10^{-3}	2.01×10^{-2}
μ_3	-	-	-	0.33
β	5	5	5	5
ω_1	0.70	0.70	0.70	0.70
ω_2	0.17	0.79	1.41×10^{-2}	3.79×10^{-2}

Note: μ are the fixed effects, ω the standard deviation of random effects and β the effect of the treatment covariate.

3.2 Design optimization and power evaluation

We aim to propose an informative design with four measurement times identical in all subjects, denoted $\xi = \{t_1, t_2, t_3, t_4\}$, first, for each candidate model separately, then, by averaging over the four candidate models, given the parameters values ψ_m given in Table 2.

3.2.1 Choice of MC-HMC samples for FIM evaluation

To perform design optimization for the presented models, we need to compute different optimality criteria, based on the FIM evaluated by MC/HMC. Since this is a stochastic approach, we can quantify uncertainty in the computation of the determinant of the FIM, according to the MC and HMC samples. For instance, the 95% uncertainty intervals of the D-criterion can be computed from the 2.5th and 97.5th percentiles of the distributions generated using non-parametric bootstrap. According to Riviere et al.,¹³ the FIM evaluation method implemented in *MIXFIM* performs well with 5000 Monte Carlo samples and 200 HMC samples, for a longitudinal logistic model for binary response (similar to our model M1) and a rich design. To ensure that these settings are appropriate for all candidate models even for sparser design (four measurement times), the convergence of the D-criterion is studied with respect to the MC samples increasing from 50 to 10,000 (see Supplementary Material, Figure S1), with a non-optimized equispaced design $\xi_{ES} = (0, 4, 8, 12)$. Two different configurations were studied: 1000 HMC samples and 1000 burn-in vs. 200 HMC and 500 burn-in. We found that the convergence plots of D-criterion shows a similar trend between the two configurations. Moreover, 5000 MC samples seem sufficient for a good convergence of the D-criterion. We therefore set the number of MC samples to 5000 and the number of HMC samples to 200 with 500 burn-in in *MIXFIM* to evaluate the FIM and to calculate the different optimality criteria with all the designs and the models of the study.

To ensure that these settings are satisfying, convergence of the D-criterion is also verified with an optimized design (Supplementary Material, Figure S2).

3.2.2 Design optimization

As in Lestini et al.,²⁹ we assumed that t_1 and t_4 are fixed to 0 and 12 months (i.e. the end of the study), only two times t_2 and t_3 are optimized among the following set: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 months. No repetition is considered, i.e. four measurement times must be different. Combinatorial optimization is performed for different optimality criteria. This corresponds to $\binom{11}{2} = 55$ possible elementary designs to be evaluated, and the optimal one is chosen according to each considered criterion. We considered $N = 100$ subjects equally distributed in the two treatment groups for design optimization, although this number has no influence on optimal sampling time or loss of efficiency in design comparisons, but only on the specific value of relative standard errors when reported.

3.2.2.1 D-, D_S - and DD_S -optimal designs for each model. Using the D-, D_S - or DD_S -optimality, the optimal location of measurement times are obtained and compared between the four models and between these optimality criteria. The relative D-, D_S - or DD_S -efficiencies of each optimal design are also computed assuming each of the four models to evaluate the impact of model misspecification on design performance.

The D_S - and the DD_S -criteria are computed with a particular interest on the treatment effect ($\psi_{S,m} = \beta$). The preliminary step to compute compound DD_S -optimal designs is to find an appropriate α_m value for each model.

For each model, we therefore compute the DD_S -optimal designs $\Xi_{DD_S,m}$ for α_m values from 0 to 1 by increment of 0.05. Then, we compare the DD_S -optimal allocations of measurement times and the D- and D_S -efficiencies of these DD_S -optimal designs over α_m . We consider that for a DD_S -design to be satisfactory, α_m should maximize $E_{D,m} \times E_{D_S,m}$. Moreover, both the D- and D_S -efficiencies should be above 0.8 (other threshold values are not investigated in this work). The final DD_S -optimal design is computed with the retained α_m value.

3.2.2.2. Robust optimal designs. To propose a robust design accounting for model uncertainty, different compound criteria (CD-, CD_S - and CDD_S -optimality) are evaluated for a combination of the four candidate models. The same weight w_m is assigned for each model (total uncertainty) i.e. $w_m = 1/4$.

3.2.3 Expected average power using FIM

We aim to evaluate the average power of the Wald test for a design $\Xi = \{N, \xi\}$ to detect a significant treatment effect over all the candidate models varying the total number of subject from 50 to 450. We also computed the number of subjects to achieve an average power of 0.9.

For that, first, we derive for each model m the standard error (SE) of β from the expected FIM and we compute the power $\pi_m(\Xi)$ to detect the treatment effect in that model, following what has been done in Retout et al.²¹ Then, the average power $\pi_{average}(\Xi)$ over M candidate models is evaluated as

$$\pi_{average}(\Xi) = \sum_{m=1}^M w_m \times \pi_m(\Xi)$$

where w_m is the weight associated to each model m used in the computation of the optimality criteria (here $1/M$, with $M = 4$).

3.3 Results

Figure 2 displays, for each model, the efficiencies with respect to each criterion, obtained with the different combinations of the second and third samples out of four measurement times.

3.3.1 D- and D_S -optimal designs for each model

Different models lead to different optimal designs. Table 3 reports the D- and D_S -efficiencies, respectively, of the D-optimal design $\Xi_{D,m}$ and D_S -optimal design $\Xi_{D_S,m}$, when the model used for data analysis is model m or another one. For instance, the design (0, 4, 5, 12) is D-optimal for M3 but underperforms for M4 with a low D-efficiency of only 0.646, and the design (0, 1, 11, 12) is D_S -optimal for M2 but conduces to a low D_S -efficiency of only 0.389 for M3.

In this work, the D_S -optimality focuses on the estimation precision of the treatment effect β . We notice that using this criterion could lead to different optimal measurement times than those obtained with the D-criterion, as in the case of M2 or M4 (Figure 2). For M2, the allocations of informative measurement times leading to a D- and D_S -efficiency above 0.8 are very similar, and the respective optimal designs are $\xi_{D,2} = (0, 1, 8, 12)$ vs. $\xi_{D_S,2} = (0, 1, 11, 12)$. However, we notice quite different efficient designs between D- and D_S -criterion for M4 ($\xi_{D,4} = (0, 6, 11, 12)$ vs. $\xi_{D_S,4} = (0, 10, 11, 12)$). This could be explained by the shape of the model curve (Figure 1, the second measurement time should be more tardive to better distinguish the two treatment groups) and by the higher number of parameters of this model compared to others (the importance of estimating β is thus relatively reduced when using D-optimality).

3.3.2 DD_S -optimal design for each model

For each model, Figure 3 shows allocated measurement times as well as D- and D_S -efficiencies of the DD_S -optimal designs over α_m possible values varying between 0 and 1. The third measurement time tends to 11 with increasing α_m for all models. These results emphasize the necessity to get tardive measurements to better distinguish the responses between the control and the treated group, and thus to better estimate the treatment effect β . The product of D- and D_S -efficiencies is, respectively, maximal when $\alpha_1 \in [0.2, 1]$, $\alpha_2 \in [0.3, 1]$, $\alpha_3 \in [0.25, 1]$ and $\alpha_4 \in [0.6, 1]$. For these values of α_m , D- and D_S -efficiencies are over 0.8 excepted for M4. With M4, D- and D_S -efficiencies are over 0.8 when $0.5 \leq \alpha_4 \leq 0.55$. In order to conserve the same range of importance for estimating the precision of parameters between the four models, we choose 0.5 as the same α_m value with

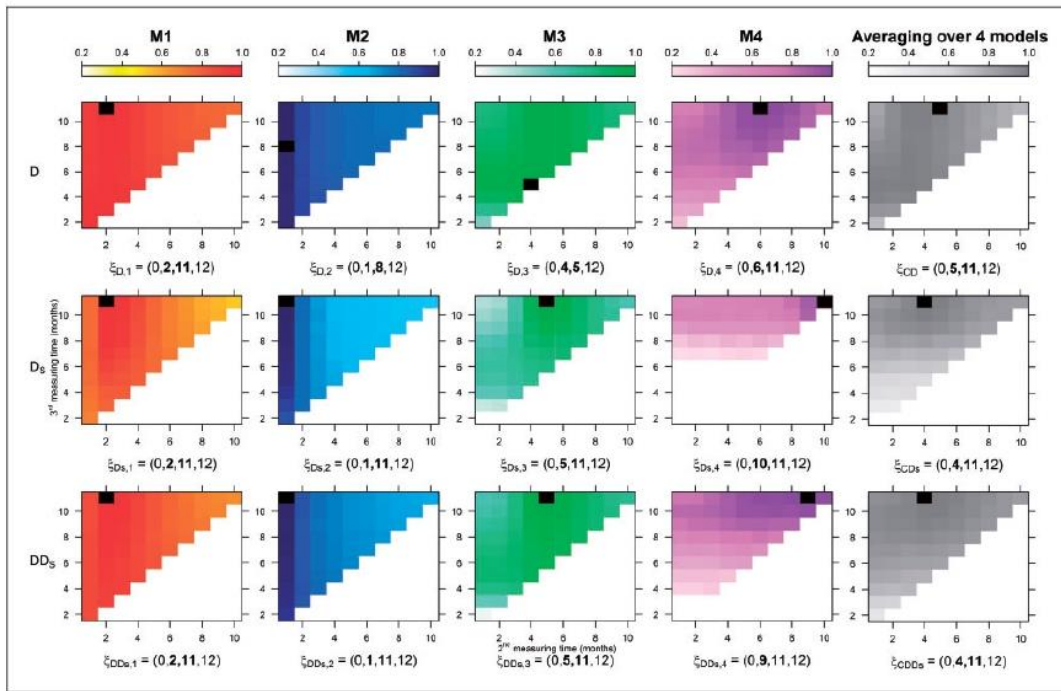


Figure 2. Heatmaps of design efficiencies, for each candidate model (M1 in red, M2 in blue, M3 in green, M4 in purple, by column) according to different considered optimality criterion (D-, D_S- or DD_S-optimality, by row). The fifth column (in grey) represents the robust criteria CD-, CD_S- or CDD_S-efficiencies. For each heatmap, the second measurement time is in abscissa and the third in ordinate. The filled black square shows the optimal design (i.e. efficiency = 1), which is also specified below each heatmap.

each model. Thus, DD_S-optimal design for each model is computed with $\alpha_m = 0.5$, with this value satisfying our two conditions.

We notice that the DD_S-criterion could lead to different optimal measurement times than those obtained with the D- or D_S-criterion (Figure 2), as in the case of M4 ($\xi_{D,4} = (0, 6, 11, 12)$, $\xi_{D_S,4} = (0, 10, 11, 12)$, $\xi_{DD_S,4} = (0, 9, 11, 12)$). These three designs lead to different predicted relative standard errors (RSE) of β , which for 100 patients, are 60%, 50%, 52%, respectively, illustrating the impact of using D_S- or DD_S-criterion on power to detect a treatment effect. For this model, we also studied the influence of the number of sampling times. We find that the DD_S-optimal design with five samples is $\xi_{DD_S} = (0, 6, 10, 11, 12)$, leading to an RSE of β of 50%, whereas the design with all 13 points lead to an RSE of 40%. For M1, M2, and M3, the DD_S-efficient designs coincide with the D_S-efficient designs. In the same way as for previous criteria (D- and D_S-efficiencies), different models lead to different DD_S-efficient designs. However, all the DD_S-optimal designs include the 11th month, which again emphasizes the importance of tardive measurement times (t_3 and t_4) in estimating the treatment effect. Table 3 reports the DD_S-efficiencies of the DD_S-optimal design $\xi_{DD_S,m}$ when the true model is m or another one. For instance, the design (0, 1, 11, 12) is DD_S-optimal for M2 but underperforms for M3 with a low DD_S-efficiency of 0.611. Therefore, this design provides poor precision on parameters of M3, especially β . With 100 individuals, the predicted RSE of β is 65% with this design vs. 41% with the DD_S-optimal for M3 ($\xi_{DD_S,3} = (0, 4, 11, 12)$).

3.3.3 Robust optimal designs

The CD-, CD_S- and CDD_S-efficiencies of every possible designs as well as the corresponding optimal design are reported in the last column of Figure 2. First, we can note that efficient robust designs are nearly the same using CD_S- or CDD_S-optimality.

Assuming model uncertainty leads to a robust CD-optimal design ($\xi_{CD} = (0, 5, 11, 12)$) which is different from the four D-optimal designs for each model. The most CD-efficient designs are closer to the D-efficient designs for the quadratic model (M3) than for other models. Table 3 presents D-efficiencies for each model. The robust CD-optimal design conducts to a D-efficiency above 0.8 regardless of the selected model. A model misspecification can lead to loss of efficiencies of up to 0.354.

Table 3. D-, D_S- and DD_S-efficiencies for different optimal designs, accounting or not for model uncertainty.

Design Ξ	$E_{D,1}$	$E_{D,2}$	$E_{D,3}$	$E_{D,4}$
$\xi_{ES} = (0, 4, 8, 12)$	0.908	0.926	0.975	0.869
$\xi_{D,1} = (0, 2, 11, 12)$	1	0.898	0.812	0.706
$\xi_{D,2} = (0, 1, 8, 12)$	0.932	1	0.875	0.787
$\xi_{D,3} = (0, 4, 5, 12)$	0.916	0.839	1	0.646
$\xi_{D,4} = (0, 6, 11, 12)$	0.829	0.802	0.960	1
$\xi_{CD} = (0, 5, 11, 12)$	0.860	0.805	0.994	0.958
$\xi_{CD_S} = (0, 4, 11, 12)$	0.910	0.828	0.977	0.882
$\xi_{CDD_S} = (0, 4, 11, 12)$	0.910	0.828	0.977	0.882
	$E_{D_S,1}$	$E_{D_S,2}$	$E_{D_S,3}$	$E_{D_S,4}$
$\xi_{ES} = (0, 4, 8, 12)$	0.841	0.646	0.835	0.393
$\xi_{D_S,1} = (0, 2, 11, 12)$	1	0.806	0.434	0.619
$\xi_{D_S,2} = (0, 1, 11, 12)$	0.751	1	0.389	0.627
$\xi_{D_S,3} = (0, 5, 11, 12)$	0.774	0.600	1	0.667
$\xi_{D_S,4} = (0, 10, 11, 12)$	0.529	0.519	0.611	1
$\xi_{CD} = (0, 5, 11, 12)$	0.774	0.600	1	0.667
$\xi_{CD_S} = (0, 4, 11, 12)$	0.865	0.637	0.906	0.655
$\xi_{CDD_S} = (0, 4, 11, 12)$	0.865	0.637	0.906	0.655
	$E_{DD_S,1}$	$E_{DD_S,2}$	$E_{DD_S,3}$	$E_{DD_S,4}$
$\xi_{ES} = (0, 4, 8, 12)$	0.882	0.754	0.924	0.723
$\xi_{DD_S,1} = (0, 2, 11, 12)$	1	0.863	0.644	0.765
$\xi_{DD_S,2} = (0, 1, 11, 12)$	0.859	1	0.611	0.755
$\xi_{DD_S,3} = (0, 5, 11, 12)$	0.827	0.722	1	0.947
$\xi_{DD_S,4} = (0, 9, 11, 12)$	0.683	0.672	0.761	1
$\xi_{CD} = (0, 5, 11, 12)$	0.827	0.722	1	0.947
$\xi_{CD_S} = (0, 4, 11, 12)$	0.893	0.751	0.954	0.894
$\xi_{CDD_S} = (0, 4, 11, 12)$	0.893	0.751	0.954	0.894

Note: D-, D_S- and DD_S-efficiencies (respectively $E_{D,m}$, $E_{D_S,m}$ and $E_{DD_S,m}$) and their corresponding compound efficiencies (E_{CD} , E_{CD_S} and E_{CDD_S}) computed for each optimal design when ignoring model uncertainty ($\xi_{D,m}$, $\xi_{D_S,m}$, $\xi_{DD_S,m}$) or accounting for model uncertainty (ξ_{CD} , ξ_{CD_S} and ξ_{CDD_S}) with each model m ($m = 1, \dots, 4$). The efficiencies are highlighted in lightgrey if between 0.5 and 0.8 and in darkgrey if below 0.5.

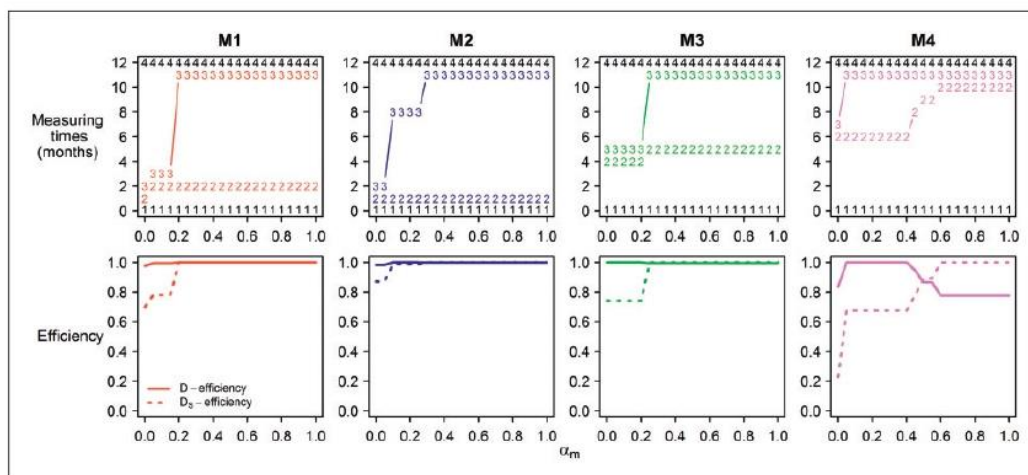


Figure 3. Results of the DD_S-optimal designs as function of the weight α_m , which quantifies the balance between the D- and the D_S-criteria, varying between 0 and 1. Results for each candidate model are represented by column (M1 in orange, M2 in blue, M3 in green, M4 in purple). First row, location of DD_S-optimal measurement times (in months): the symbol x represents the xth measurement in each design. Second row, corresponding D- (solid line) and D_S-efficiency (dotted line), with respect to the D- and D_S-optimal design, respectively.

In the same way as using CD-optimality, assuming model uncertainty leads to CD_S -optimal design close to D_S -optimal design for model M3: $\xi_{CD_S} = (0, 4, 11, 12)$ and $\xi_{D_S, 3} = (0, 5, 11, 12)$. Furthermore, CD_S -efficient designs include mostly more tardive measurement times than the CD-efficient designs (Figure 2). The robust CD_S -optimal design also conducts to a D-efficiency above 0.8 regardless of the candidate model and reduces in average the maximal loss of D_S -efficiency as compared to the D_S -optimal designs obtained separately for each model (Table 3, 0.363 vs. respectively 0.566, 0.619, 0.4 and 0.481 with $\xi_{D_S, 1}$, $\xi_{D_S, 2}$, $\xi_{D_S, 3}$ and $\xi_{D_S, 4}$). The CD_S -optimal design could be particularly appropriate to precisely estimate β , especially if the final model is M3 (D_S -efficiency of 0.906).

Assuming model uncertainty leads to a robust CDD_S -optimal design ($\xi_{CDD_S} = (0, 4, 11, 12)$) which is different from the four local DD_S -optimal designs for each model. The robust CDD_S -optimal design is the same as the CD_S - and thus conducts to a D-efficiency above 0.8 regardless of the candidate model. It also reduces the maximal loss of DD_S -efficiency as compared to the DD_S -optimal design obtained separately for each model (Table 3, 0.249 vs. respectively 0.356, 0.389, 0.278 and 0.328 with $\xi_{DD_S, 1}$, $\xi_{DD_S, 2}$, $\xi_{DD_S, 3}$ and $\xi_{DD_S, 4}$), and shows DD_S -efficiencies above 0.8 for M1, M3 and M4.

3.3.4 Expected average power using FIM

Expected power over number of subjects N , under each model and in average, are presented with different designs in Figure 4: the CDD_S -optimal or robust design ($\xi_{CDD_S} = (0, 4, 11, 12)$), the DD_S -optimal for M1 ($\xi_{DD_S, 1} = (0, 2, 11, 12)$) and the non-optimized equispaced design ($\xi_{ES} = (0, 4, 8, 12)$). The exponential model (M4) presents the worst expected power among the four candidate models, especially with ξ_{ES} , mainly because this model involves one more estimated parameter than the three other models. Among these three designs, the ξ_{CDD_S} presents the closest curves of power under the four models. Moreover, to reach a $\pi_{average}$ of 0.9, the predicted number of subjects needed (NSN) is 274 with ξ_{CDD_S} vs. 320 with $\xi_{DD_S, 1}$ and 358 with ξ_{ES} . In our example, the CDD_S -criterion is able to propose design giving decent performances in terms of predicted power whatever the model chosen among the candidates.

4 Evaluation by simulation

In this section, we perform simulations to assess the relevance of the design strategy presented above. For this purpose, with simulated datasets of 274 subjects (i.e. the NSN to reach an average power of 0.9 with the robust design ξ_{CDD_S}), we (1) compare performances of ξ_{CDD_S} vs. a locally optimal design for the linear model M1 $\xi_{DD_S, 1}$ and a non-optimized equispaced design ξ_{ES} in terms of bias and precision of estimates and (2) evaluate the type I error and power as well as the adequacy between predictions and simulation results with the robust design.

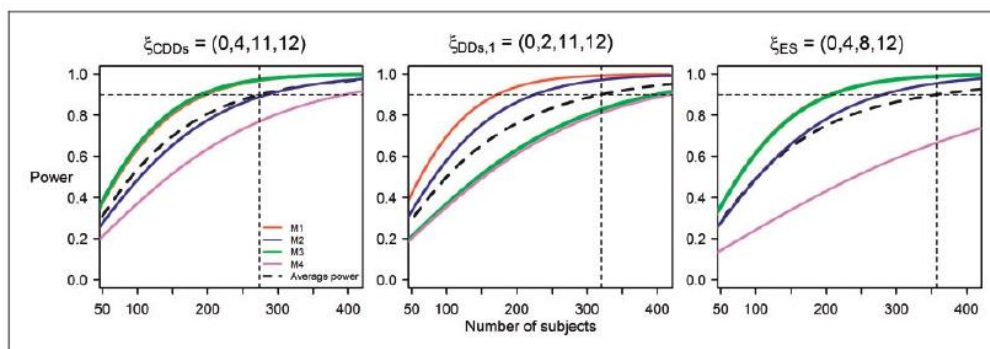


Figure 4. Predicted power for each model and in average with different designs, for varying total number of subjects from 50 to 450. From the left to the right, with the robust design ξ_{CDD_S} , the design optimized for model M1 $\xi_{DD_S, 1}$ and the non-optimized design ξ_{ES} , the expected power from FIM for varying number of subjects is displayed in orange for M1, blue for M2, green for M3 and purple for M4. The average power over the four models is in dashed line. The dotted line crossing the x-axis shows the number of individuals needed to obtain an average power of 0.9.

4.1 CTS

4.1.1 Designs and models

Repeated binary response trials of 274 individuals (137 per treatment group) are simulated under each model m (M1 to M4) with their respective population parameters values indicated in Table 2.

For the first objective (to compare different designs performances), with ξ_{CDD_s} , $\xi_{\text{DD}_{s,1}}$ and ξ_{ES} , $K_{1,m} = 500$ datasets are simulated with each model under H_1 (treatment effect $\beta = 5$) and each one of these three designs.

To achieve the second objective, with the robust design ξ_{CDD_s} , $K_{0,m} = 500$ additional datasets are simulated with each model under H_0 (without treatment effect, $\beta = 0$), in order to evaluate the type I error.

4.1.2 Evaluation

For each dataset, population parameters are estimated by the SAEM algorithm³² implemented in MONOLIX 2016R1 (www.lixoft.eu), a software devoted to maximum likelihood estimation of parameters in NLMEMs.

First, from the analysis of the $K_{1,m}$ simulated datasets, we calculate the inverse of the variance-covariance matrix computed from 500 estimated parameter vectors, for each of the three designs. We defined the observed relative D-efficiency between designs as the ratio of determinants of these matrices, normalized by the number of parameters. We also compare the distribution of the relative estimation error of parameters for each model (associated with the relative bias values) between the three designs.

Then, with the design ξ_{CDD_s} , the adequacy between the FIM predictions and the CTS observations is evaluated (in terms of parameter precision and power). The empirical standard errors (SE_{CTS}) of each parameter of each model are defined as the standard deviation of all estimated values. The empirical RSEs are then calculated as the ratio of the SE_{CTS} to the parameter simulated values. The 95% confidence intervals for each empirical RSE are given as $\left[\sqrt{\frac{q_1 \times \text{RSE}^2}{K_{1,m} - 1}}; \sqrt{\frac{q_2 \times \text{RSE}^2}{K_{1,m} - 1}} \right]$, with q_1 and q_2 , respectively, the 2.5% and 97.5% quantiles of the χ^2 distribution with $K_{1,m} - 1 = 499$ degrees of freedom. To evaluate the relevance of MC/HMC-based approach to evaluate the expected FIM (see Section 2.1.3.) and the parameter estimation method, predicted RSEs obtained using the expected FIM are compared with empirical RSE and relative root mean square error (RRMSE) obtained with CTS. Thus, the type I error and the power are evaluated. From the $K_{0,m}$ datasets simulated with $\beta = 0$, the type I error of the Wald test on the treatment effect is evaluated as the proportion of trials for which H_0 is rejected and compared with 0.05. The 95% prediction interval (95% PI) for proportion π of H_0 rejection (type I error or power) is computed by binomial test.³³ As explained in Section 3.2.1., we can evaluate the uncertainty in the computation of the D-criterion and the expected power derived from FIM, according to the MC and HMC samples. Under H_1 , we can thus give a confidence interval as 2.5th and 97.5th percentiles of bootstrap obtained with our FIM computation approach, around the FIM expected power value. The observed proportion of significant Wald tests is compared with the FIM expected power value and its (respectively prediction or confidence) intervals, by binomial test or by bootstrap.

4.2 Results

4.2.1 Comparison of robust design with non-optimized and non-robust design

Informativeness of the three designs is investigated here. For each model, ratios of observed normalized determinant of the variance-covariance matrix of these designs are computed with the robust design ξ_{CDD_s} as reference.

As presented in Table 4, overall, the estimations of parameters are more precise with the robust design than with the equispaced design for all the models (observed D-efficiency with respect to ξ_{CDD_s} is below 1). With the optimal design for M1 $\xi_{\text{DD}_{s,1}}$, the parameters of M1 and M2 are slightly more precise than with the robust design, but this design induces important loss of precision for the parameters of M3 and M4 compared to the robust design.

The three designs show acceptable and similar distribution of relative estimation errors (REE, Figure 5) under M1, M2 and M3. REE under M4 are larger than other models with the three designs. The robust design performs better on M4 than the other designs, especially for the parameters μ_2 and ω_2 which present larger range of whiskers. Moreover, the maximal relative bias with ξ_{CDD_s} is 15% but is over 20% with $\xi_{\text{DD}_{s,1}}$ (37% for μ_2 and 35% for ω_2 on M4) and ξ_{ES} (47% for μ_2 and 35% for ω_2 on M4).

Table 4. Observed D-efficiency computed from clinical trial simulations for each model, with respect to the robust design.

	M1	M2	M3	M4
ξ_{CDDs} : reference	/	/	/	/
$\xi_{DDs,1}$	1.11	1.06	0.79	0.74
ξ_{ES}	0.98	0.94	0.98	0.84

Note: For each model M1 to M4, the inverse of the variance-covariance matrix of parameter estimates is calculated for each of the three designs: the DD_s-optimal design for M1 ($\xi_{DDs,1}$), the non-optimized equispaced design (ξ_{ES}) and the robust design (ξ_{CDDs}). The observed D-efficiency between designs is defined as the ratio of determinants of these matrices, normalized by the number of parameters.

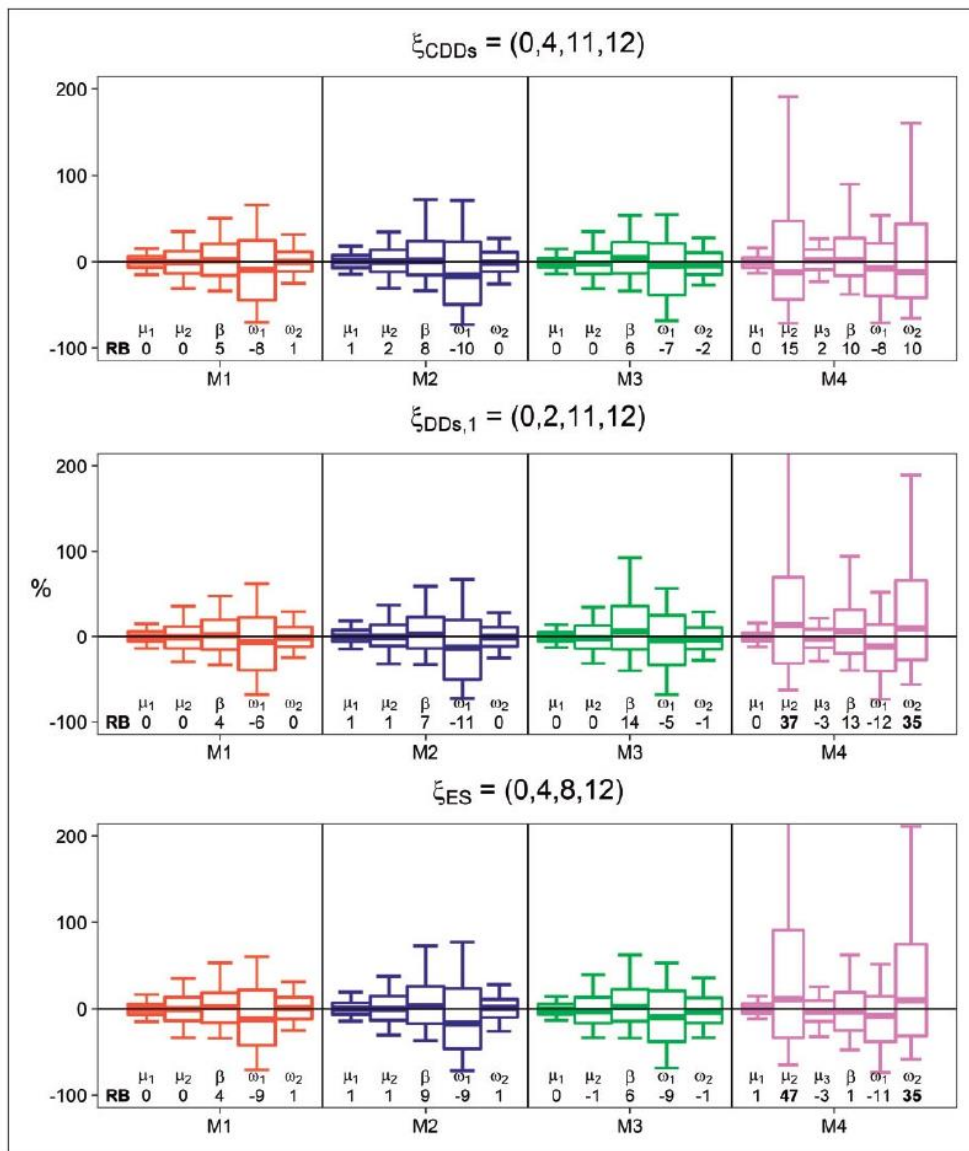


Figure 5. Boxplot of the relative estimation error (REE, in %, whiskers from 5th to 95th percentile) for each parameter of the four models: M1 in orange, M2 in blue, M3 in green and M4 in purple. From the top to the bottom with 500 datasets and 274 subjects by dataset, the robust design ξ_{CDDs} , the optimal design of M1 $\xi_{DDs,1}$ and the equispaced design ξ_{ES} . The relative bias (RB, in %) of each parameter is positioned under each corresponding boxplot, the values in bold font highlight the RB over 20%. μ are the fixed effects, β the covariate effect and ω the standard deviation of random effects.

4.2.2 Adequation between FIM predictions and CTS results with the robust design

As shown in Figure 6 for all models, the predicted RSE by the FIM computation (RSE_{FIM}) are in the same range as empirical RSE from CTS (RSE_{CTS}) especially for the fixed effect parameters (μ) and for standard deviations of random effects parameters (ω), even when the model is more complex (M4, excepted μ_3). RSE of the covariate effect β are, however, slightly under-predicted. Nevertheless, the maximal differences between RSE_{FIM} and RSE_{CTS} are at most 4%. Moreover, the population parameters of all candidate models could be estimated with reasonable estimation error for the proposed robust design (RSE around 30% for the fixed effects and 50% for the random effects) except M4 which contains six parameters: θ_2 is quite difficult to be estimated with such a sparse design.

Table 5 displays the type I error of the Wald test on the treatment effect in each candidate model, using the estimated or empirical SE. For the Wald tests using estimated SE, it can be seen that the type I error lies in the 95% PI of the nominal level [0.033,0.073] for all models. Table 5 shows also the predicted power from FIM vs. observed power from CTS under H_1 . Two different kinds of intervals are built around the FIM predicted power: as under H_0 using the binomial test or as percentiles of bootstrap performed with FIM evaluation. We can note that the interval built by bootstrap is smaller than the binomial test one. The choice of MC and HMC samples seems therefore appropriate for power predictions (see convergence plots in Supplementary Material, Figure S2). Nevertheless, for each model, the observed power is slightly above the predicted power. These results are partly due to the positive bias on β (Figure 5) and developed further in the discussion.

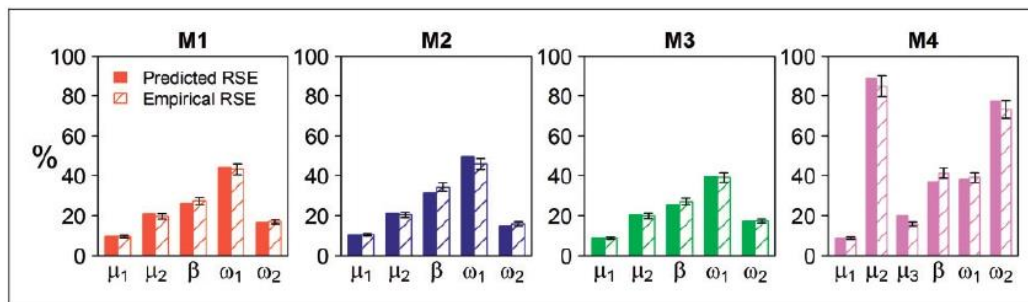


Figure 6. Comparison of predicted and observed estimation errors on all parameters of the four candidate models with 274 subjects and the robust design ξ_{CDDs} . Relative standard error RSE (%) for the candidate models M1 to M4 are predicted using the expected FIM (full bar). Empirical RSE (hatched bar) are obtained by clinical trial simulations (CTS) with 500 datasets. μ are the fixed effects, β the covariate effect and ω the standard deviation of random effects.

Table 5. Wald tests predictions and simulations on the covariate effect, under H_0 and under H_1 .

	Under H_0		Under H_1			
	Nominal type I error	Observed type I error	FIM predicted power		Observed power	
			Binomial PI_{95}	Bootstrap CI_{95}		
M1		0.048	0.967	[0.949, 0.982]	[0.963, 0.971]	0.988
M2	0.05	0.060	0.889	[0.859, 0.916]	[0.873, 0.902]	0.988
M3	[0.033,0.073]	0.068	0.973	[0.959, 0.986]	[0.969, 0.976]	0.996
M4		0.036	0.769	[0.731, 0.806]	[0.741, 0.792]	0.860

Note: Under H_0 , the 95% prediction interval calculated by binomial test is [0.033,0.073] for 500 simulations. For each model M1 to M4, the predicted power to detect a covariate effect β under H_1 using the Fisher information matrix (FIM) is calculated with the design $\Xi = \{N = 274, \xi_{CDDs} = (0.4, 11, 12)\}$. Two intervals are presented around the predicted power: a prediction interval obtained by binomial test (PI) and an uncertainty interval (CI) using bootstrap from MC/HMC method for FIM evaluation. The observed power from clinical trial simulations (CTS) is given as the observed proportion of significant Wald tests performed on the covariate effect β with estimated standard error from 500 simulated clinical trials with a type I error of 5%.

5 Application to a real study

5.1 Presentation of the study

We applied the proposed approach to a clinical study conducted by the National Institute on Aging³⁴ using the published model, i.e. not assuming model uncertainty. The study EPESE (Established Populations for Epidemiologic Studies of the Elderly) was a longitudinal study with 4162 persons aged 65 years or older involved. The outcome of interest was the instrumental activities of daily living (IADL). The IADL was measured once a year during three years. The evolution of IADL was compared between those who were cognitively impaired at baseline and those who were cognitively intact.

In Landerman et al.,³⁵ the IADL tasks (traveling, shopping, preparing meals, doing housework and managing finances without assistance) were dichotomized: individuals unable to perform four or five tasks were classified as disabled ($y_{ij}=1$) and those able to perform at least two tasks were not classified as disabled ($y_{ij}=0$). This dichotomous outcome was measured at 0, 1, 2 and 3 years. They studied the cognitive impairment at baseline on the outcome. The proposed model in Landerman et al.³⁵ to describe probability to be disabled at time t_{ij} was

$$\text{logit}(p(y_{ij} = 1|b_i)) = \theta_1 + \beta_1 \times 1_{BI} + (\theta_2 + \beta_2 \times 1_{BI})t_{ij}$$

where 1_{BI} is the cognitive impairment at baseline ($1_{BI} = 0$ if no impairment, $1_{BI} = 1$ if impairment). β_1 was the impairment effect on the intercept parameter θ_1 and β_2 the impairment effect on the slope parameter θ_2 . θ_1 follows a normal distribution: $\theta_1 = \mu_1 + b_1$, where $b_1 \sim N(0, \omega_1^2)$. No interindividual variability was considered on θ_2 ($\theta_2 = \mu_2$).

5.2 Design evaluation and optimization

5.2.1 Methods

To evaluate the relevance of the FIM-based approach, using the EPESE design $\xi_{\text{EPESE}} = (0, 1, 2, 3)$ and the provided model described above, we compared the SE predicted using MC/HMC (see Section 2.1.3.) and the SE reported in Landerman et al.³⁵ after data fitting.

Then, based on the DD_S-criterion (with the subset of parameters of interests ψ_S composed of μ_2 and β_2), we propose an optimized design of four measuring times, assuming that t_1 and t_4 are fixed to 0 and 3 years (i.e. the end of the study), and that only two times t_2 and t_3 are optimized among 0.2 to 2.8 years by increment of 0.2 years. No repetition is considered, i.e. four measurement times must be different. Combinatorial optimization is performed for different optimality criteria. This corresponds to $\binom{14}{2} = 91$ possible elementary designs to be evaluated, and the optimal one is chosen according to each considered criterion.

Finally, we computed the power of the Wald test and the NSN is calculated (as explained in Section 3.2.3.) for the effect of cognitive impairment on the slope, β_2 , for the slope parameter μ_2 .

5.2.2 Results

The comparison of the observed and the FIM computed SE of the parameters is reported in Table 6. SE of each fixed or random effect is close between FIM predictions and results reported in Landerman et al.³⁵

Table 6. Parameter estimation precision in EPESE analysis (Observed) in Landerman et al.³⁵ and FIM evaluation (Predicted).

	Parameter values	Observed SE	Predicted SE
μ_1	-7.205	0.372	0.270
μ_2	1.677	0.107	0.074
β_1	5.604	0.384	0.367
β_2	0.034	0.140	0.157
ω_1^2	15.017	1.757	1.316

Note: μ are the fixed effects, ω the standard deviation of random effect and β the effect of the cognitive impairment.

According to the D-criterion, the D-optimal design with the provided model is $\xi_D = (0, 1.2, 2.8, 3)$. The evolution of the probability to be disabled and the cognitive impairment effect on this evolution were particularly of interest in this study. Thus, we compute the D_S -criterion with μ_2 and β_2 in the subset of parameters ψ_S (see equation (5) in Table 1). The D_S -optimal design is $\xi_{D_S} = (0, 1.2, 2.8, 3)$, i.e. the same measuring times allocation of the D-optimal design. Moreover, using the DD_S -criterion with α between 0.4 (i.e. D-criterion) and 1 (i.e. D_S -criterion) leads to the same optimal design ($\xi_{DD_S} = (0, 1.2, 2.8, 3)$). As for the linear model M1 of Section 3, all the possible designs lead to D-efficiencies higher than 0.8. For instance, with the ξ_{EPESE} , the D-efficiency is 0.944.

The provided model in Landerman et al.³⁵ shows that the slope μ_2 is non null. Moreover, only 31 individuals would have been needed with ξ_{DD_S} (vs. 33 with ξ_{EPESE}) to reach a power of 0.8. However, the covariate β_2 is not significant. With the ξ_{EPESE} , the NSN to detect this covariate with a power of 0.8 would be 484,518 subjects. With the optimized design $\xi_{DD_S} = (0, 1.2, 2.8, 3)$, it would be 437,087 subjects. Recruiting this number of individuals, even with the optimized design, would be not feasible. Nevertheless, this result highlights that the DD_S -optimal design can reduce the NSN compared to a non-optimized design.

6 Discussion

In this paper, we propose a new methodological strategy based on MC/HMC and compound optimality theory to design longitudinal trials with discrete outcomes. Our approach accounts for model uncertainty when assuming a set of candidate models and ensures a compromise between the overall precision of estimation and the power of the Wald test to detect a covariate effect. To our knowledge, this is the first time that a robust optimal design approach, based on MC/HMC to compute the FIM without any linearization, is evaluated by extensive CTS. The relevance of using the FIM to efficiently predict the power of the Wald test has been evidenced before but only in the context of repeated continuous outcomes.^{21,36} We have now confirmed the utility of this FIM-based method to predict average power across several candidate models, with an example of repeated discrete outcomes including two treatment groups. Although the considered example is quite theoretical here (Sections 3 and 4), the proposed method is illustrated with a published study analyzing real data (Section 5). In this study, the repeated binary outcome was the ability to undertake life tasks (based on the Instrumental Activities Daily Living tasks). The chosen settings can also easily be extended to clinical longitudinal studies with binary outcomes such as remission of dyskinesia in Parkinson's Disease. In such cases, it is not possible to obtain dense measuring times because medical tests to evaluate this scale are very time-consuming. In another context, a method to design longitudinal studies was proposed when dynamics is modelled by a binary Markov process, using examples of infection by *Streptococcus pneumoniae*.³⁷

To determine informative designs, we study different optimality criteria, according to different purposes of the study. The D-optimality is used to optimize the precision of the whole set of population parameters. The D_S -optimality allows to accommodate situations in which only a subset of the model parameters is of interest. It is particularly useful to minimize the standard error on the treatment effect, thus to maximize the power of the Wald test of the study and therefore to reduce the number of subjects needed. A parameter which is a primary end-point (i.e. treatment effect) or a key secondary end-point (i.e. time effect) of a study should be included in the subset of parameters of interest. This subset could also be defined based on the future use of the model. For example, in CTS, if outcomes are sensitive to some specific parameters, they should be included in the subset of parameters of interest. However, this criterion usually leads to designs that do not ensure the experimental identifiability.²² To overcome this issue, the DD_S -optimality allows the experimenter to find a compromise between the D- and the D_S -criterion,²⁰ by weighting each one according to the importance given to each model parameter. The weight for each criterion depends of the objective of the study, and its meaning is further discussed in Cook and Wong.³⁸ In spite of its usefulness, the DD_S -optimality was mostly used with continuous standard nonlinear regression models^{39,40} but has not been reported so far in the context of discrete NLMEMs, to our knowledge. Another novelty of this work is to extend these standard optimal criteria to account for model uncertainty, using their weighted product to propose compound CD-, CD_S - and CDD_S -criterion over a set of candidate models.

In the motivation example, we considered several candidate binary NLMEMs to describe the logit-probability of response over time: linear (M1), loglinear (M2), quadratic (M3) and exponential (M4). M4 contains one more parameter than other models and therefore provides information about the predicting abilities of the MC/HMC method with a more complex model. The non-optimized equispaced design in our example provides D-efficiencies > 0.8 under each candidate model. This design shows, however, very poor D_S -efficiency of 0.393 under M4 (Table 3). When optimizing designs, we find that accounting or not for uncertainty in models and

focusing or not on the estimation precision of the covariate effect may lead to different optimal locations of measurement times. Moreover, misspecification of models in case of local design optimization (i.e. assuming a given model) can lead to D-, D_S - or DD_S -efficiency, respectively, as low as 0.666, 0.389 or 0.611 for some models. The proposed compound CD-, CD_S - or CDD_S -optimal designs provides a better compromise for different candidate models and greatly reduces the loss of efficiency. These three compound optimal designs conduce all to D-efficiencies of at least 0.8 for each model. Furthermore, in terms of ability to detect the treatment effect, choosing the robust design ξ_{CDD_S} instead of the design optimal for the linear model $\xi_{DD_S,1}$ or either a non-optimized ξ_{ES} reduces the required sample size of the study. The NSN for an average power of 0.9 is indeed 274 with ξ_{CDD_S} vs. 320 with $\xi_{DD_S,1}$ and 358 with ξ_{ES} . The NSN is, however, close with either the CDD_S - or the CD-design (274 vs. 280). This could be explained by the fact that these two designs only differ by their respective second time point (respectively, at fourth and fifth month). However, the gain in power could be larger in other examples, with more differences in allocation of sampling times between the CDD_S - and the CD-design. The three optimal designs using compound criteria CD-, CD_S - or CDD_S are close, which is partly explained by the use of a combinatorial optimization, with only one possible measurement per month, limiting the total number of possible designs. Surprisingly, the second measurement time t_2 of the CD-design is latter than the one of the CD_S -design, whereas the CD_S -design maximizes the average power. It is explained by the loglinear model (M2) for which early measurement seems already informative to distinguish the logit probability of response between the two treatment groups. Overall, we notice that with this set of models, picking only tardive measurement could be very suboptimal to maximize the precision of the parameter β and thus the power to detect a treatment effect. The closeness of D- and D_S -optimal measurement times under each model and averaging over the four models (using CD- and CD_S -optimality) indicates that a precise estimation of all the parameters (and probably, even more the “trend” parameters μ_2 and ω_2 on which β acts) is necessary to reach a satisfactory power. In the optimization procedure, it would also be interesting to consider reducing the duration of the study.

We also evaluate our approach by simulating clinical trials under each candidate model. To assess the relevance of the robust design, we compare performances between three different designs: a non-optimized design with a measurement time every four months, a design optimal for a linear model (M1, maximizing the DD_S -optimality) and a robust design accounting for model uncertainty (maximizing the compound DD_S -optimality). In our example, choosing the robust design avoids important loss of information on parameters, especially when the right model was exponential (M4). Other scenarios could have been examined, e.g. simulating data under another model M5 and fit under each candidate model M1 to M4. We also could evaluate the power performances of model averaging²³ vs. model selection as in Buatois et al.,²⁸ but it is not the purpose of the present work. Furthermore, a method has been recently proposed to optimize designs when the analysis is done by model averaging.⁴¹

With the CDD_S -design and 274 individuals, in spite of a slight under or over-prediction of RSE for some parameters, the discrepancy is at most 4%; thus, the prediction using the FIM evaluated by MC/HMC is accurate. This approach avoids extensive CTS and can efficiently help to easily detect non-informative designs in case of large predicted RSE values. Using the predicted SE given by FIM, we predict the power of the Wald test to show significant difference between treatment groups. Under H_0 , we note a good control of the type I error by CTS, despite the asymptotic properties of the Wald test (explained in Dubois et al.⁴²) and our sparse measurement design. Under H_1 , we found differences in predicted power between the different models: respectively 0.967, 0.889, 0.973 and 0.769 under M1, M2, M3 and M4. Whereas the SE for treatment effect (β) are not over-predicted, powers are under-predicted under all the models M1 to M4 (especially M2 and M4): observed power are, respectively, 0.988, 0.988, 0.996 and 0.86. However, we note that the power is never over-predicted. Moreover, the FIM predicted Wald statistic is close to the median of the observed Wald statistic distribution, even with M2 and M4 (Supplementary Material, Figure S2). We investigated the possibility that the number of MC/HMC samples would be insufficient under our design to adequately predict the precision on parameter β (and thus of the power). However, the convergence plots of the determinant of the FIM and of the SE on β showed acceptable results (Figure S3, Supplementary Material). Discrepancies between predicted and observed power are mainly explained by the positive bias on β in CTS. If β would be estimated at its true simulated value (5), the estimated power would be close to the predicted values (Supplementary Material, Table S1). Other settings of the SAEM algorithm of MONOLIX were investigated, but none provided more satisfactory relative bias. To our knowledge, an evaluation of different algorithms for repeated binary responses has never been published, as it has been done for ordered categorical⁴³ or continuous data.⁴⁴ The analysis of this kind of data is challenging, and requires informative design to avoid problem of bias. Further research works on the estimation and design method for this kind of data are necessary. It could thus lead to fair measures of discrepancies between FIM prediction and CTS observations, and also to a fair comparison of different designs in terms of average power.

We apply the described FIM based methods, using a published logistic model built on real data.³⁵ This model is describing the evolution of the disability probability over time, including parameters as the baseline (fixed and random effect), the time effect or the cognitive impairment effect. With the model and the design elements provided in the study, we find that the SE predicted by FIM are close to those provided in the study, showing that the FIM evaluation method by MC/HMC is relevant. The provided model is the only one mixed model which is reported. If other possible models would have been proposed, we would be able to compute compound criteria, with each model weight depending on its likelihood, in accordance with the model averaging principle.²³ Therefore, a robust design based on the CDD_S-criterion would have been proposed. However, with the given model, we propose to optimize the choice of the measuring times according to the DD_S-criterion. We also predict the number of subjects that would be needed to reach a power of 0.8 to detect the time effect and the cognitive impairment. We show that, instead of using a non-optimized equispaced design, optimizing the design using the DD_S-criterion can reduce the number of subjects needed.

In this work, we perform combinatorial optimization by evaluating FIM for every possible design, an alternative approach would consist in implementing an optimization algorithm to improve computing time as suggested in Riviere et al.¹³ More efficient algorithms could also be explored, such as stochastic approximation, annealing methods, randomized exchange algorithm,⁴⁵ multiplicative method⁴⁶ or particle swarm optimization.^{47,48} It is also important to account for uncertainty in parameters in addition to uncertainty in models,^{49,50} by evaluating the expectation of the FIM over the distribution of population parameters instead of assuming known values of these parameters. We also plan to integrate these robust design methods (over the parameter distribution and over the models) in a next version of PFIM, which is the software program for designing longitudinal studies developed by our team (www.pfim.biostat.fr).⁵¹ Furthermore, robust design methods studied in this paper could be combined with adaptive designs,^{52,53} by using the accumulated information at each stage of the study to update knowledge of the candidate models and of the parameter distributions to be taken into account in design optimization.

The proposed design strategy based on the expected FIM evaluated by MC/HMC and the compound optimality theory is a relevant approach, which enables, for the first time, optimization of informative sparse designs for longitudinal binary data, over several candidate models. This approach can also be applied to other type of discrete data models, such as multi-category or Poisson count data.⁵⁴ A robust design method and application using *MIXFIM* for Poisson models was studied in Loingeville et al.⁵⁰ but not for the comparison of two treatment groups, hence using CD-optimality. The approach described in this article also aims to propose design, providing good power to detect the treatment effect and acceptable precision of all parameters while accounting for model uncertainty.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article

ORCID iD

Jérémy Seurat  <https://orcid.org/0000-0002-8127-7751>

Supplemental material

Supplemental material for this article is available online.

References

1. Lalonde RL, Kowalski KG, Hutmacher MM, et al. Model-based drug development. *Clin Pharmacol Ther* 2007; **82**: 21–32.
2. Mould DR and Upton RN. Basic concepts in population modeling, simulation, and model-based drug development. *CPT Pharmacometrics Syst Pharmacol* 2012; **1**: e6.

3. Mould DR and Upton RN. Basic concepts in population modeling, simulation, and model-based drug development – part 2: introduction to pharmacokinetic modeling methods. *CPT Pharmacomet Syst Pharmacol* 2013; **2**: e38.
4. Upton RN and Mould DR. Basic concepts in population modeling, simulation, and model-based drug development: part 3 – introduction to pharmacodynamic modeling methods. *CPT Pharmacomet Syst Pharmacol* 2014; **3**: e88.
5. Syrjänen RK, Kilpi TM, Kaijalainen TH, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Finnish children younger than 2 years old. *J Infect Dis* 2001; **184**: 451–459.
6. Mestre TA, Beaulieu-Boire I, Aquino CC, et al. What is a clinically important change in the Unified Dyskinesia Rating Scale in Parkinson's disease? *Parkinsonism Relat Disord* 2015; **21**: 1349–1354.
7. Atkinson A, Donev A and Tobias R. *Optimum experimental designs, with SAS*. Oxford, New York: Oxford University Press, 2007.
8. Pronzato L and Pázman A. *Design of experiments in nonlinear models: asymptotic normality, optimality criteria and small-sample properties*. New York: Springer Science & Business Media, 2013.
9. Fedorov V and Leonov S. *Optimal design for nonlinear response models*. Boca Raton: CRC Press, 2013, pp.20–29.
10. Mentré F, Mallet A and Baccar D. Optimal design in random-effects regression models. *Biometrika* 1997; **84**: 429–442.
11. Nyberg J, Bazzoli C, Ogungbenro K, et al. Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies. *Br J Clin Pharmacol* 2015; **79**: 6–17.
12. Nguyen TT and Mentré F. Evaluation of the Fisher information matrix in nonlinear mixed effect models using adaptive Gaussian quadrature. *Comput Stat Data Anal* 2014; **80**: 57–69.
13. Riviere M-K, Ueckert S and Mentré F. An MCMC method for the evaluation of the Fisher information matrix for nonlinear mixed effect models. *Biostatistics* 2016; **17**: 737–750.
14. Mielke T. *Approximation of the Fisher information and design in nonlinear mixed effects models*. PhD Thesis. Otto-von-Guericke-Universität Magdeburg, Germany, 2012.
15. Waite TW and Woods DC. Designs for generalized linear models with random block effects via information matrix approximations. *Biometrika* 2015; **102**: 677–693.
16. Waite TW. *Design of experiments with mixed effects and discrete responses plus related topics*. PhD Thesis. University of Southampton, United Kingdom, 2012.
17. Ueckert S and Mentré F. A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature. *Comput Stat Data Anal* 2017; **111**: 203–219.
18. Riviere M-K and Mentré F. R package MIXFIM, version 1.0 <http://mc-stan.org/> (accessed 17 August 2015).
19. Stan Development Team. Rstan: the R interface to Stan, version 2.12.0. <http://mc-stan.org/> (accessed 19 January 2018).
20. Atkinson AC and Bogacka B. Compound D- and Ds-optimum designs for determining the order of a chemical reaction. *Technometrics* 1997; **39**: 347–356.
21. Retout S, Comets E, Samson A, et al. Design in nonlinear mixed effects models: optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates. *Stat Med* 2007; **26**: 5162–5179.
22. Sölkner J. Choice of optimality criteria for the design of crossbreeding experiments. *J Anim Sci* 1993; **71**: 2867–2873.
23. Buckland ST, Burnham KP and Augustin NH. Model selection: an integral part of inference. *Biometrics* 1997; **53**: 603–618.
24. Atkinson AC. DT-optimum designs for model discrimination and parameter estimation. *J Stat Plan Infer* 2008; **138**: 56–64.
25. Nguyen TT, Bénech H, Delaforge M, et al. Design optimisation for pharmacokinetic modeling of a cocktail of phenotyping drugs. *Pharm Stat* 2016; **15**: 165–177.
26. Schorning K, Bornkamp B, Bretz F, et al. Model selection versus model averaging in dose finding studies. *Stat Med* 2016; **35**: 4021–4040.
27. Aoki Y, Röshammar D, Hamrén B, et al. Model selection and averaging of nonlinear mixed-effect models for robust phase III dose selection. *J Pharmacokinet Pharmacodyn* 2017; **44**: 581–597.
28. Buatois S, Ueckert S, Frey N, et al. Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. *AAPS J* 2018; **20**: 56.
29. Lestini G, Ueckert S and Mentré F. Model-based optimal robust design in pharmacometrics. PODE, Uppsala, Sweden, 2016. www.maths.qmul.ac.uk/~bb/PODE/PODE2016_JLestini.pdf (accessed 30 May 2017).
30. Kiefer J and Wolfowitz J. The equivalence of two extremum problems. *Can J Math* 1960; **12**: 363–366.
31. Ogungbenro K and Aarons L. Population Fisher information matrix and optimal design of discrete data responses in population pharmacodynamic experiments. *J Pharmacokinet Pharmacodyn* 2011; **38**: 449–469.
32. Kuhn E and Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal* 2005; **49**: 1020–1038.
33. Clopper CJ and Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; **26**: 404–413.
34. Blazer D, Burchett B, Service C, et al. The association of age and depression among the elderly: an epidemiologic exploration. *J Gerontol* 1991; **46**: M210–M215.
35. Landerman LR, Mustillo SA and Land KC. Modeling repeated measures of dichotomous data: testing whether the within-person trajectory of change varies across levels of between-person factors. *Soc Sci Res* 2011; **40**: 1456–1464.
36. Nguyen TT, Bazzoli C and Mentré F. Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models. *Stat Med* 2012; **31**: 1043–1058.

37. Mehtälä J, Auranen K and Kulathinal S. Optimal designs for epidemiologic longitudinal studies with binary outcomes. *Stat Methods Med Res* 2015; **24**: 803–818.
38. Cook RD and Wong WK. On the equivalence of constrained and compound optimal designs. *J Am Stat Assoc* 1994; **89**: 687–692.
39. Yeatts SD, Gennings C and Crofton KM. Optimal design for the precise estimation of an interaction threshold: the impact of exposure to a mixture of 18 polyhalogenated aromatic hydrocarbons. *Risk Anal* 2012; **32**: 1784–1797.
40. Winkens B, Schouten HJA, van Breukelen GJP, et al. Optimal designs for clinical trials with second-order polynomial treatment effects. *Stat Methods Med Res* 2007; **16**: 523–537.
41. Alhorn K, Schorning K and Dette H. Optimal designs for frequentist model averaging. *arXiv:180705234 [stat]*, <http://arxiv.org/abs/1807.05234> (2018, accessed 1 August 2018).
42. Dubois A, Lavielle M, Gsteiger S, et al. Model-based analyses of bioequivalence crossover trials using the stochastic approximation expectation maximisation algorithm. *Stat Med* 2011; **30**: 2582–2600.
43. Savic RM, Mentré F and Lavielle M. Implementation and evaluation of the SAEM algorithm for longitudinal ordered categorical data with an illustration in pharmacokinetics–pharmacodynamics. *AAPS J* 2010; **13**: 44–53.
44. Plan EL, Maloney A, Mentré F, et al. Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose-response models. *AAPS J* 2012; **14**: 420–432.
45. Harman R, Filová L and Richtárik P. A randomized exchange algorithm for computing optimal approximate designs of experiments. *Journal of the American Statistical Association*. In press. <https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1546588>.
46. Yu Y. Monotonic convergence of a general algorithm for computing optimal designs. *Ann Statist* 2010; **38**: 1593–1606.
47. Wong WK, Chen R-B, Huang C-C, et al. A modified particle swarm optimization technique for finding optimal designs for mixture models. *PloS One* 2015; **10**: e0124720.
48. Kim S and Wong WK. Extended two-stage adaptive designs with three target responses for phase II clinical trials. *Stat Methods Med Res* 2018; **27**: 3628–3642.
49. Foo LK, McGree J, Eccleston J, et al. Comparison of robust criteria for D-optimal designs. *J Biopharm Stat* 2012; **22**: 1193–1205.
50. Loingeville F, Nguyen TT, Riviere M-K, et al. Robust designs in longitudinal studies accounting for parameter and model uncertainties - application to count data. *J Biopharm Stat*. In press. <https://www.tandfonline.com/doi/abs/10.1080/10543406.2019.1607367?journalCode=lbps20>.
51. Bazzoli C, Retout S and Mentré F. Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *Comput Methods Programs Biomed* 2010; **98**: 55–65.
52. Lestini G, Dumont C and Mentré F. Influence of the size of cohorts in adaptive design for nonlinear mixed effects models: an evaluation by simulation for a pharmacokinetic and pharmacodynamic model for a biomarker in oncology. *Pharm Res* 2015; **32**: 3159–3169.
53. Strömberg EA and Hooker AC. The effect of using a robust optimality criterion in model based adaptive optimization. *J Pharmacokinetic Pharmacodyn* 2017; **44**: 317–324.
54. Fang X, Li J, Wong WK, et al. Detecting the violation of variance homogeneity in mixed models. *Stat Methods Med Res* 2016; **25**: 2506–2520.

Supplementary Material

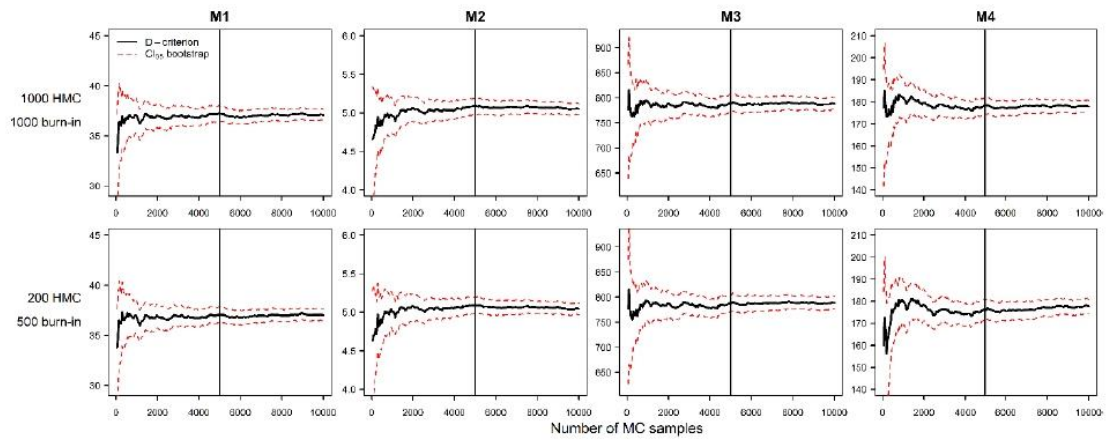


Figure S1. Convergence plots of D-criterion with a non-optimized design. The D-criterion is plotted in black as a function of the number of Monte Carlo samples, with the equispaced design, for each model M1 to M4 by column. For the computation of the Fisher information matrix, two different algorithm settings are displayed by row (1000 Hamiltonian Monte-Carlo (HMC) samples with 1000 burn-in on the left or 200 HMC with 500 burn-in on the right). The red dotted lines represent the 2.5th and the 97.5th percentiles of the D-optimality values using bootstrap.

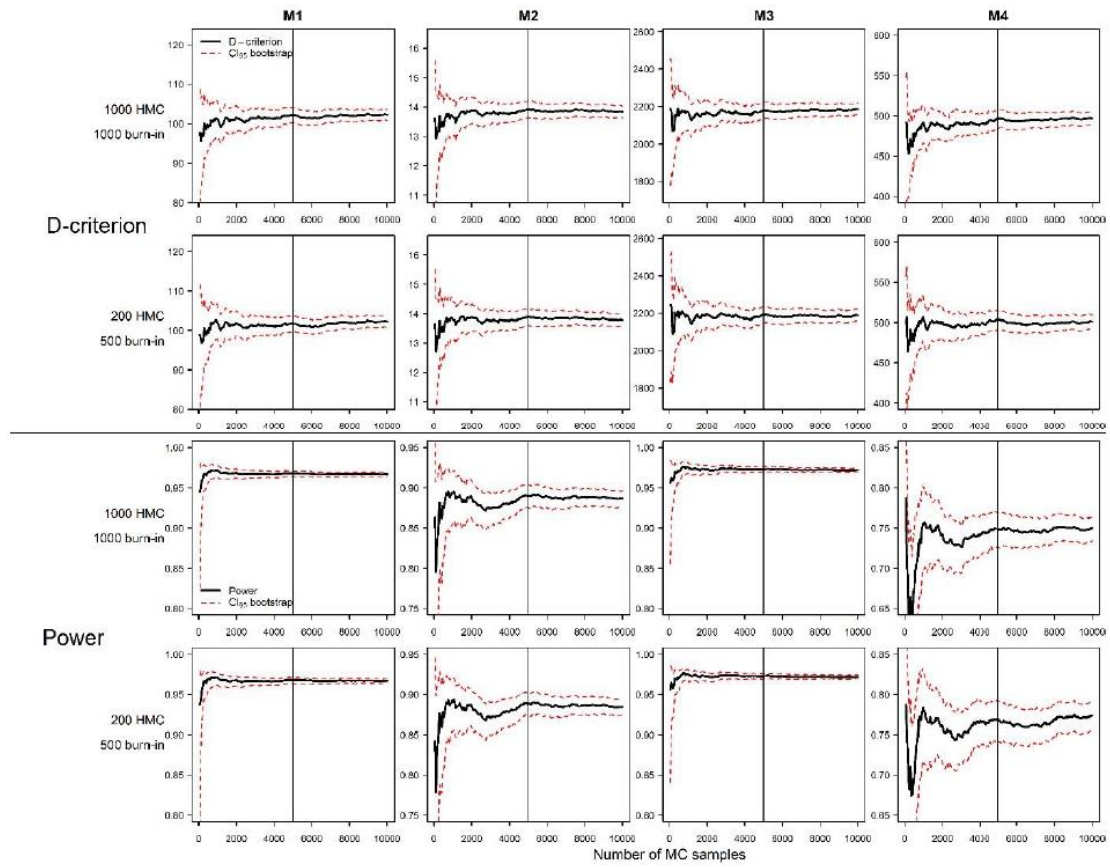


Figure S2. Convergence plots of D-criterion and power with the robust design. For each model M1 to M4 by column, the D-criterion on the top part and the power on the bottom part, are plotted in black as a function of the number of Monte Carlo samples with the robust design. Different algorithm settings for the computation of the Fisher information matrix are represented (1000 Hamiltonian Monte-Carlo (HMC) samples and 1000 burn-in on the first row of each part or 200 HMC and 500 burn-in on the second row of each part). The red dotted lines represent the 2.5th and the 97.5th percentiles of the bootstrap values

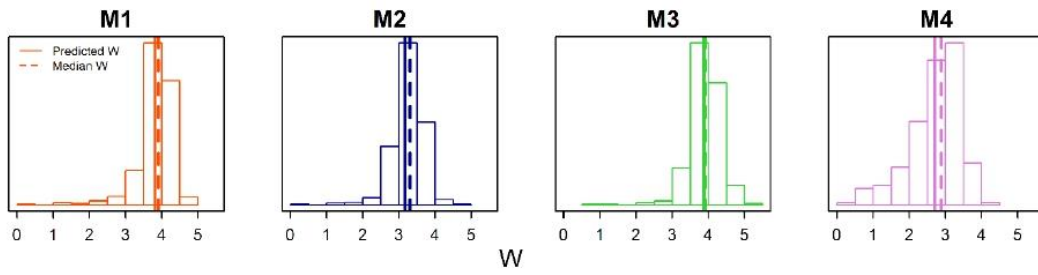


Figure S3. Distribution of observed Wald test statistic with the robust design. For each model M1 to M4, the distribution of the 500 observed Wald test statistic W is represented as histograms. For each model, the solid line is the FIM predicted W and the dashed line is the median of observed W .

Table S1. Observed power of the Wald test without bias on β with the robust design.

	M1	M2	M3	M4
FIM predicted power	0.968	0.889	0.973	0.77
CTS observed power	0.988	0.988	0.996	0.86
CTS observed power using simulated	0.976	0.914	0.992	0.746

FIM predicted power are obtained using the true value of β (5) and the SE() calculated from FIM evaluation. CTS observed power are obtained using estimated β and their SE() as given by the SAEM algorithm. CTS observed power using simulated β are obtained using the true value of β (5) and SE() as given by the SAEM algorithm.

2.2. Algorithmes pour l'optimisation de protocoles

2.2.1. Algorithmes multiplicatifs

2.2.1.1. Résumé

Nous avons proposé d'étendre un algorithme dans le but d'optimiser des protocoles d'études longitudinales, en utilisant la matrice d'information de Fisher. Nous nous sommes intéressés aux algorithmes multiplicatifs, conçus pour l'optimisation de protocoles dans un espace de design discret. Les algorithmes multiplicatifs ont été principalement utilisés pour les régressions individuelles. Dans ce travail, ils ont été étendus aux modèles mixtes et ont été codés dans deux langages : R et C.

Contrairement à l'utilisation de l'algorithme de Fedorov-Wynn, l'utilisateur n'a pas besoin de définir un protocole pour l'initialisation de cet algorithme multiplicatif. Toutefois, il est nécessaire que la matrice de Fisher soit calculée pour chaque protocole élémentaire possible ξ_q , avec $q = (1, \dots, Q)$. Les ξ_q sont définis à partir de contraintes (par exemple, chaque ξ_q peut être défini par 1 dose parmi 10 possibles et un vecteur de 5 temps de mesures parmi 8 possibles, soit un nombre total de protocoles élémentaires $Q = 10 \times 56 = 560$). A l'initialisation de l'algorithme, le même poids (c'est à dire le même nombre de sujets N_q) est attribué à chaque ξ_q . Ensuite, à partir du critère D, les poids optimaux de chaque ξ_q sont calculés par l'algorithme suivant un processus itératif. La monotonie et la convergence de l'algorithme ont été prouvées à partir d'une précédente démonstration de la convergence des algorithmes multiplicatifs (Yu 2010).

Deux exemples illustratifs d'optimisation de protocole d'études longitudinales ont été choisis afin d'appliquer l'algorithme développé. Le premier exemple traite d'une optimisation de temps de mesures et/ou de choix de doses pour une étude longitudinale sur la dégénérescence maculaire, en utilisant un modèle caractérisant l'évolution de l'acuité visuelle, définie comme une valeur continue (Rosenfeld et al. 2006; Buatois et al. 2018). La FIM a été évaluée par méthode FO. Deux cas d'optimisation de bras de doses parmi 7 possibles (et du nombre d'individus dans chaque bras) ont été réalisés, avec respectivement 26 et 4 temps de mesures fixés. L'optimisation des doses (et du nombre de sujets à attribuer à chaque dose) par l'algorithme a conduit à un gain de D-efficacité de respectivement 13 et 14 % par rapport aux protocoles de référence respectifs (non optimisés, avec patients équitablement répartis dans 4

bras de doses). Un exemple plus complexe a été étudié, avec recherche à la fois des doses et des temps de mesures les plus informatifs (1932 protocoles élémentaires possibles). Le protocole de population optimisé était composé de 4 groupes avec des doses et temps de mesures différents. Ce protocole optimisé conduisait à un gain d'efficacité de 26%. Cet exemple a permis aussi de comparer la vitesse des versions codées (en C et en R) de l'algorithme. L'algorithme en C était 14 fois plus rapide que l'algorithme en R pour cet exemple (31 s vs. 432 s, respectivement). Le deuxième exemple étudié, aussi utilisé pour appliquer la méthode de planification de protocoles robustes (section 2.1.), consiste en une optimisation de temps de mesures (2 parmi 4) pour des données binaires analysées par modèle logistique, avec 55 protocoles élémentaires possibles. L'algorithme a permis de trouver un protocole présentant un gain de D-efficacité de 18 % par rapport au protocole de référence, non optimisé (avec le même intervalle de temps entre deux mesures).

Ces travaux ont montré que l'utilisation d'un algorithme multiplicatif est une solution efficace pour l'optimisation de protocoles d'études longitudinales. Ils ont fait l'objet d'un article en révision dans *Computer Methods and Programs in Biomedicine*.

L'algorithme présenté a été intégré à la nouvelle version du programme PFIM (section 2.2.2.).

2.2.1.2. Article 2 (accepté)

Finding optimal design in nonlinear mixed effect models using multiplicative algorithms, publié dans *Computer Methods and Programs in Biomedicine*, 2021

Jérémy Seurat*, Yuxin Tang*, France Mentré, Thu Thuy Nguyen

* : co-premier auteur

Finding optimal design in nonlinear mixed effect models using multiplicative algorithms

Jérémy Seurat*, Yuxin Tang, France Mentré, Thu Thuy Nguyen

Université de Paris, INSERM, IAME, F-75006 Paris, France

Abstract

Background and objectives: To optimize designs for longitudinal studies analyzed by nonlinear mixed effect models (NLMEMs), the Fisher information matrix (FIM) can be used. In this work, we focused on the multiplicative algorithms, previously applied in standard individual regression, to find optimal designs for NLMEMs. **Methods:** We extended multiplicative algorithms to mixed models and implemented the algorithm both in R and in C. Then, we applied the algorithm to find D-optimal designs in two longitudinal data examples, one with continuous and one with binary outcome. **Results:** For these examples, the algorithm performed faster in C than in R. Design optimization using the multiplicative algorithm led to designs with D-efficiency gains between 13% and 25% compared to non-optimized designs. **Conclusion:** We found that the multiplicative algorithm can be used efficiently to design longitudinal studies.

▀

Keywords: nonlinear mixed effect model, Fisher information matrix, multiplicative algorithm, optimal design, D-optimality, longitudinal studies

*Corresponding author

Email address: jeremy.seurat@inserm.fr (Thu Thuy Nguyen)

The first two authors equally contributed.

1. Introduction

Nonlinear mixed effect models (NLMEMs) are frequently used in model-based drug development to analyze all longitudinal data obtained during clinical trials, notably pharmacokinetic/pharmacodynamic data [1, 2, 3, 4]. NLMEMs can compensate for the lack of individual information by borrowing the strength from the longitudinal data in the whole population and therefore allow for precise parameter estimation even with sparse designs, where few samples are collected from each subject, according to experimental constraints. Nevertheless, it is crucial to find informative experimental designs before modeling to get a precise estimation of parameters [5]. Indeed, the informativeness of a dataset for parameter estimation depends on the design elements as the number of subjects, the number and time of the measurements or the doses.

To evaluate and compare designs, the theory of optimum experimental design in classic nonlinear models was introduced [6, 7, 8], based on the expected Fisher information matrix (FIM), as an alternative to time-consuming clinical trial simulation. The inverse of the FIM, according to the Cramer-Rao inequality, is the lower bound of the variance covariance matrix of any unbiased estimators of the parameters. Then, the predicted standard errors for estimated parameters can be calculated, as the square roots of the diagonal elements of the inverse of the FIM.

Once the FIM is evaluated, criteria based on FIM can be computed to optimize designs. The D-optimality criterion is widely used for design optimization. It consists in maximizing the determinant of the FIM, normalized by the number of parameters to be estimated. Several algorithms have been developed for design optimization [6]. Particularly in PFIM (www.pfim.biostat.fr), the first R program dedicated to design evaluation and optimization of longitudinal studies [9], Simplex [10] and Fedorov-Wynn [11, 12, 13] algorithms were implemented. The Simplex algorithm optimized designs within a continuous design space while the Fedorov-Wynn algorithm works in a discrete design space. Each type of algorithm corresponds to a different optimization strategy. The Simplex

algorithm can be useful to explore all the design space, but could result in time points that have no clinical relevance. The Fedorov-Wynn type algorithm is one of the earliest and most notable algorithms for finding optimal designs in a discrete space and can be proved to converge to the optimal design for convex criterion functions. However, it does not have good computation speed in complex design problems [14].

With the increasing use of modeling techniques when analyzing longitudinal data, there is a need to explore new optimization algorithms to design these studies appropriately. Multiplicative algorithms optimize in discrete design space and their convergence for D-optimality and other criteria has been generally established [15]. Several advantages were highlighted with multiplicative algorithms, in particular a better computation speed than the Fedorov-Wynn algorithm [14]. So far, multiplicative algorithms have been mostly explored in standard individual regression. That is why we aim to explore these algorithms further in mixed effect models and to illustrate it with different applications.

In this article, we present the principles of a multiplicative algorithm [15] and how it can be applied to optimize population designs in clinical trials analyzed by NLMEMs. The presented optimization method is based on expression of the FIM. We illustrate these methods through two longitudinal data examples. The first example is inspired by a longitudinal study of a monoclonal antibody indicated in the treatment of wet age-related macular degeneration (wet-AMD) [16]. Considering a mixed effect model previously developed for a continuous outcome, we aim to optimize both the measurement times and doses in patients. The second example is a trial with repeated binary outcomes in two treatment groups, modeled by a mixed logistic model, inspired by a previous study [17, 18]. We aim to optimize the measurement times considering either identical design for the two treatment arms.

In section 2, we introduce the notations for NLMEMs and FIM. In section 3, we present the multiplicative algorithm that we used in this work. Sections 4 show an application of the algorithm. Finally, we discuss the results and perspectives of this work in section 5.

2. Population Fisher Information Matrix

2.1. Design

Let N be the number of subjects and ξ_i the elementary design of individual i , defined by the number n_i of samples and their allocation in time $(t_{i1}, \dots, t_{in_i})$, and, in some cases, additional design variables (like dose d_i). The design space for ξ_i is discrete *i.e.* ξ_i is included in the set of ξ_q with $q = 1, \dots, Q$. The Q possible elementary designs are predefined according to the constraints that are given in terms of dose and time measurements (number and allocation). Usually, the population design Ξ is defined by the total number of subjects N and the Q elementary designs with weight w_q associated with each design ξ_q as: $\Xi_w = \begin{pmatrix} \xi_1 & \dots & \xi_Q \\ w_1 & \dots & w_Q \end{pmatrix}$, so that $\Xi = (N, \Xi_w)$, with the weight set $w = (w_1, \dots, w_Q)$. Each weight represents the proportion of the total number of subjects N to be taken for each elementary design ξ_q (the number of subjects allocated to one elementary design is $N_q = N \times w_q$). Each w_q satisfies the equation $\sum_{q=1}^Q w_q = 1$, $0 \leq w_q \leq 1$, $\forall q = 1, \dots, Q$. The algorithm tries to find the optimal weight set w , *i.e.* the optimal proportions of subjects allocated to each possible elementary design ξ_q .

In practice, after optimization of the population design (*i.e.* after optimization of the w_q by the multiplicative algorithm), designs with a very low weight (*i.e.* $N \times w_q = 0$ after rounding to the closest integer) were discarded. Then, weights corresponding to the Q' remaining designs are summed up and each remaining weight is divided by the sum to obtain new weights $w_{q'}$ whose sum is equal to 1. The number of subjects per group $N_{q'}$ is rounded to the closest integer of $N \times w_{q'}$, as done when using the Fedorov-Wynn algorithm for nonlinear mixed models [13], with $N_1, \dots, N_{Q'}$ satisfying the equation $\sum_{q'=1}^{Q'} N_{q'} = N$.

2.2. Nonlinear mixed effect models for continuous data

The NLMEM to describe observations y_i (of length n_i) of the i^{th} individual, with its design ξ_i included in the set of $(\xi_q)_{q=1,\dots,Q}$ has the following form:

$$y_i = f(\theta_i, \xi_i) + \varepsilon_i \quad (1)$$

where the function f provides a n_i -vector of model predicted values for individual i and its elementary design ξ_i . An example of f is given in section 4.2 (with f a disease progression model for visual acuity). θ_i is the vector of the parameters for individual i , expressed as a function $g(\mu, b_i, z_i, \beta)$ of: the vector of fixed effects μ , the vector of random effects b_i and the vector of covariates z_i for individual i , as well as the vector of covariate effects β . Random effects b_i follow a normal distribution with mean zero and variance-covariance matrix Ω which accounts for the inter-individual variability, *i.e.* $b_i \sim \mathcal{N}(0, \Omega)$. Ω_u is the vector containing all unique elements of Ω . Assuming a simple model without covariate effects, g is usually additive, so that $\theta_i = \mu + b_i$, or exponential so that $\theta_i = \mu \times e^{b_i}$. $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T$ is the vector of random errors for individual i and n_i is the number of its observations. $\varepsilon_i \sim \mathcal{N}(0, \Sigma_i)$, with Σ_i a diagonal matrix of dimension n_i (assuming independent ε_i for each subject) defined as:

$$\Sigma_i = \text{diag}(\sigma_{inter} + \sigma_{slope} f(\theta_i, \xi_i))^2 \quad (2)$$

where σ_{inter} and σ_{slope} are standard deviations of additive (*i.e.* the constant part) and proportional terms of the error model, respectively.

The vector Ψ^T of length P contains all population parameters to be estimated ($\Psi^T = (\mu^T, \beta^T, \Omega_u^T, \sigma_{inter}, \sigma_{slope})$).

The notations of NLMEMs for discrete data are available in Appendix A.

2.3. Population Fisher information matrix

Assuming independence across N individuals, the population FIM, $M_{PF}(\Psi, \Xi_w)$, for a population design Ξ_w , is defined by the weighted sum of all the elementary matrices $M_{PF}(\Psi, \xi_q)$,

$$M_{PF}(\Psi, \Xi_w) = \sum_{q=1}^Q w_q M_{PF}(\Psi, \xi_q) \quad (3)$$

For individual i , the elementary FIM, $M_{PF}(\Psi, \xi_i)$, for the elementary design ξ_i , is defined as

$$M_{PF}(\Psi, \xi_i) = E_{y_i, z_i} \left(\frac{\partial \log L(\Psi, y_i, z_i)}{\partial \Psi} \frac{\partial \log L(\Psi, y_i, z_i)^T}{\partial \Psi} \right) \quad (4)$$

where $\log L(\Psi, y_i, z_i)$ is the log-likelihood of the vector of observations of individual i for parameter Ψ , given by

$$\log L(\Psi, y_i, z_i) = \log \left(\int_{b_i} p(y_i | b_i, z_i, \Psi) p(b_i | z_i, \Psi) db_i \right) \quad (5)$$

with $p(y_i | b_i, z_i, \Psi)$ the probability density function (p.d.f.) of y_i given parameters Ψ , random effects b_i and covariates z_i , and with $p(b_i | z_i, \Psi)$ the p.d.f. of b_i given Ψ and z_i .

As there is no analytic expression for the likelihood in NLMEMs, approximations such as linearization can be used to approximate the Fisher information matrix [19, 20]. For continuous outcome models, when linearizing the model around a guess value of the fixed effects, the variance of the observations is then independent of the fixed effects, which leads to a block-diagonal expression of the FIM [20]. It has been shown that this expression of the FIM performed better than the full matrix expression given by other linearization approaches, providing results closer to those obtained by clinical trial simulation [21]. This block diagonal expression, computed by default in PFIM program v4.0, is the one used to evaluate the FIM with continuous outcomes in this work (section 4). The method used in this work to evaluate the FIM for discrete outcome models is presented in Appendix A.

The D-optimality criterion ϕ_D is a widely used optimality criterion in the field of optimal design, which consists in maximizing the determinant of the FIM $M_{PF}(\Psi, \Xi_w)$ (see equation (3)), normalized by the number of model parameters to be estimated P :

$$\phi_D(\Xi_w) = |M_{PF}(\Psi, \Xi_w)|^{\frac{1}{P}} \quad (6)$$

The relative D-efficiency $E_D(\Xi_w)$ of a design Ξ_w , with respect to the reference design Ξ_w^* , is computed as:

$$E_D(\Xi_w) = \frac{\phi_D(\Xi_w)}{\phi_D(\Xi_w^*)} \quad (7)$$

3. Multiplicative algorithm for design optimization

The monotonic convergence of this class of multiplicative algorithm has generally been established under certain conditions by several optimality criteria, including the D-criterion [15]. In this paper, the proposed algorithm is an extension [15, 22] to NLMEMs, based on the D-optimality criterion. Moreover, this class of algorithm has been shown to converge for several applications, including design optimization with pre-specified possible design points [22], as it is the case in this manuscript.

3.1. Equivalence theorem

Optimizing a population design Ξ_w means optimizing the weight set w , *i.e* the weight w_q of each elementary design ξ_q , knowing that all the possible elementary designs are defined beforehand. The equivalence theorem can be used to assess whether any given design is optimal for a given criterion [23]. The general equivalence theorem [24] states that the following conditions for a D-optimal design are equivalent, assuming the concavity of the D-criterion $\phi_D(\Xi_w)$ as formulated by Yu [15]:

1. Ξ_w^* is the D-optimal design
2. Ξ_w^* maximizes $\phi_D(\Xi_w) = |M_{PF}(\Psi, \Xi_w)|^{\frac{1}{P}}$
3. Ξ_w^* minimizes $\max_{1 \leq q \leq Q} (d(w_q))$, $\forall q = 1, \dots, Q$, where $d(w_q)$ is the maximal rate of increase of ϕ_D at w_q :

$$d(w_q) = \frac{\partial \phi_D(\Xi_w)}{\partial w_q} = \text{tr} \left(\frac{1}{P} |M_{PF}(\Psi, \Xi_w)|^{\frac{1}{P}-1} M_{PF}(\Psi, \Xi_w)^{-1} M_{PF}(\Psi, \xi_q) \right) \quad (8)$$

3.2. Multiplicative algorithm and convergence criterion

The algorithm is defined as follows. The vector of weights $w = (w_1, \dots, w_Q)$ is initialized as $w^{(0)}$. At each iteration k , each element of this vector is multiplied by a factor, depending on the results of the previous iteration:

$$w_q^{(k+1)} = m_q^{(k)} w_q^{(k)} \quad (9)$$

Iterations process until convergence to the optimal design or until the maximum number of iteration K is reached.

The multiplier factor at each iteration k is defined as:

$$m_q^{(k)} = \frac{d(w_q^{(k)})^\lambda}{\sum_{j=1}^Q w_j^{(k)} d(w_j^{(k)})^\lambda} \quad (10)$$

where $\lambda \in (0, 1]$, denotes a calibration parameter (in case of D-optimality, the convergence can be shown if $\lambda = 1$ [25]). This parameter is set at 0.99 in our examples.

The general equivalence theorem ensures the monotonicity of the multiplicative algorithm to a D-optimal design. Further proofs can be found in the article of Yu, especially theorem for monotonicity and for convergence [15]. Notice that at this step, for the algorithm itself the space of interest is the weight set w . Each element w_q of this weight set corresponds to each possible elementary design ξ_q given before performing the algorithm. As D-criterion ϕ_D is increasing and concave, Yu proposed to use the following convergence criterion: $w^{(k)}$ at the k^{th} iteration is considered optimal if each weight w_q satisfies the inequality:

$$\max_{1 \leq q \leq Q} d(w_q^{(k)}) \leq (1 + \delta) \sum_{j=1}^Q w_j^{(k)} d(w_j^{(k)}) \quad (11)$$

where δ is a small positive constant ($\delta = 10^{-4}$ according to Yu [15]).

The D-optimal design Ξ_w^* is found if at certain iteration, the weight set w satisfies the equation (11).

Heuristic explanation

In order to maximize the determinant of the moment matrix, the algorithm proposes to find each optimal weight w_q (corresponding to each pre-defined elementary design ξ_q) by multiplying it by the factor m_q at each iteration, which depends on $d(w_q)$. The value of $d(w_q)$ indicates the amount of gain in information, as measured by ϕ_D , by a slight increase in w_q , the weight on the q^{th} elementary design. Then, the equation (9) can be seen as adjusting the

weight vector so that relatively more weight is placed on elementary designs whose increased weight may result in a larger gain in ϕ_D .

3.3. Implementation

Two versions of multiplicative algorithm implementation are proposed. One is implemented in an R function, the other is implemented in C (see Appendices B and C, respectively) for the loop part of the algorithm which needs to be compiled as a dynamic link library file and then to be called by R. Given that the exact CPU time varies with each time of execution and depends on the computer's hardware, values provided are meant to compare relatively with each other. Arguments of the function are presented in Figure 1 and commented codes are available in Appendices B and C. In order to account for several methods of FIM evaluation, the list of evaluated FIMs is considered as an argument of the function. Each FIM in that list quantifies the information of each elementary design generated from the design space given experimental constraints. The codes for FIM evaluation using the FO method and for the algorithm application corresponding to the illustrative example of section 4 are available in Appendices D and E, respectively. The presented function returns the weight vector at each iteration (in the form of a dynamic graph), the number of iterations required, whether the convergence of the algorithm is achieved, the optimal design and the D-criterion value corresponding to the final weight vector.

4. Application to design optimization for a longitudinal study with continuous outcome

4.1. Context and objectives

This example is inspired by the clinical development of a monoclonal antibody indicated in the treatment of wet age-related macular degeneration (wet-AMD) [16]. The treatment reduces vessel leakiness and improves visual acuity (VA) by neutralizing the vascular endothelial growth factor (VEGF) in the

Function in R

- `multiplicative_algo(FIMs, version, designSpace, nbTot, lambda, delta, iteration)`

The arguments for this function represent:

<i>FIMs</i>	the list of elementary FIM corresponding to all possible elementary designs according to given constraints
<i>version</i>	the version of implementation for the loop in the algorithm, either R or C
<i>designSpace</i>	the list or vector of elementary design space corresponding to the list of FIMs
<i>nbTot</i>	total number of subject considered in the study
<i>lambda</i>	lambda's value in algorithm
<i>delta</i>	the constant's value for stop condition
<i>iteration</i>	maximum number of iterations per call of function

justification=raggedright,singlelinecheck=false

Figure 1: Arguments of the multiplicative algorithm function in R.

retina. In wet-AMD, disease progression and treatment effect are assessed using a VA test based on the early treatment diabetic retinopathy study (ETDRS) chart which contains 14 lines (70 letters). The ETDRS chart is used to measure the number of letters successfully read by the patient at a given visit. The trial duration is set to 24 months, with several VA measurements per patient. The patients were randomized to different dose groups.

Both VA measurement times and doses were optimized in these patients for a hypothetical future study, based on a published model of disease progression and the anti-VEGF treatment effect, which have been used in a previous methodological study [26].

4.2. Model and parameters

The following model f describes the time course of VA (a continuous scale from 0 to 70 letters read by the patient) in wet-AMD patients, with or without anti-VEGF treatment, where f provides a predicted scalar value at time t :

$$f(\theta, d, t) = \text{VA}_0 + (1 - e^{-k t}) \times \left(\frac{E_{max} d}{(ED_{50} + d)} - \beta \text{VA}_0 \right) \quad (12)$$

Parameter vector θ is composed of 5 elements: $\theta = (\text{VA}_0, k, \beta, E_{max}, ED_{50})$. VA_0 is the visual acuity at baseline and in the absence of treatment. VA exponentially decreases over time t in untreated patients, reaching an asymptote $\text{VA}_0 \times (1 - \beta)$ at a rate k . Based on k , one can derive the average time to steady

state $(5 \times \log(2)/k)$, which equals approximately 24 months and corresponds to the end of trial. The model mimics a mixture of both disease-modifying and symptomatic effects. The dose-response relationship is assumed to be an Emax function, where E_{max} represents the maximum number of letters that an individual can gain and ED_{50} the dose at which 50% of the maximal effect is achieved. For a very high dose and at steady state, the predicted visual acuity is $VA_0 \times (1 - \beta) + E_{max}$.

The E_{max} parameter follows a normal distribution, VA_0 , k , and β follow a log-normal distribution. No inter-individual variability of ED_{50} is considered. An additive residual error model is assumed (section 2.2). The parameter values are reported in Table 1.

Parameter	distribution	μ	ω^2	σ^2
VA_0 (Letters)	log-normal	55	0.07	-
k (Day ⁻¹)	log-normal	0.005	0.5	-
β	log-normal	0.2	1.0	-
E_{max} (Letters)	normal	30	150	-
ED_{50} (μg)	log-normal	150	-	-
σ^2 (Letters ²)	-	-	-	28

Table 1: Parameter values of the model for visual acuity (with μ the fixed effects, ω^2 the variance of the random effects, and σ^2 the variance of the residual error).

The predicted VA time course curves from 0 to 24 months with four dose groups (0, 50, 100 and 500 μg) are shown in Figure 2 for illustration of the model shape, using the fixed effects values of Table 1.

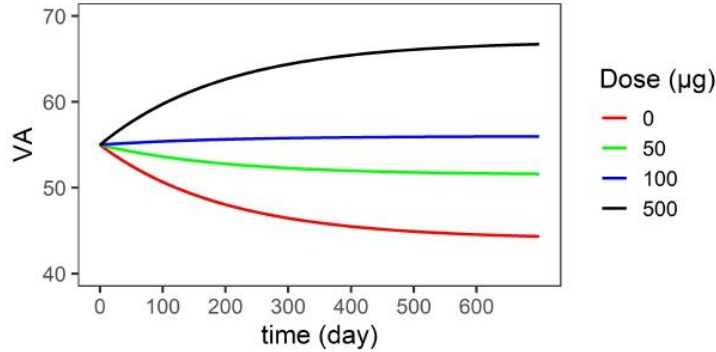


Figure 2: Median time course of the visual acuity (VA, in *Letters*) predicted by the model [\(12\)](#), for different dose groups: 0, 50, 100 and 500 μg respectively.

4.3. Design optimization

The multiplicative algorithm was used to optimize the population design Ξ_w , based on the D-optimality criterion. The population FIM was evaluated using the first-order approximation of the model [\(9\)](#).

The population design Ξ was composed of $N = 300$ patients included in total and assigned to different elementary designs ξ_q , chosen among a finite number Q of possible elementary designs. The measurement times were optimized among 26 possible times: at baseline (day 0, fixed), day 7 and every 28 days from the 28th to the 672nd day. The optimal doses were chosen from among 7 possible doses: 0, 25, 50, 100, 150, 300 and 500 μg . Given these constraints, each elementary design ξ_q was composed of a vector t_q of several measurement times and one dose d_q : $\xi_q = (t_q, d_q)$.

First, a reference-rich design was evaluated with all 26 measurement times in 4 dose groups (0, 150, 300, 500 μg) defined in a previous study [\(26\)](#) and with the same number of subjects in each group ($\xi_q = (t_q = \{0, 7, \text{seq}(28, 672, 28)\}, d_q)$). In other words, this population design is composed of 4 elementary designs with the same "weight" allocated to each design (Table [2](#) Reference 26 times scenario).

This reference design was compared to the one with optimized dose groups

Scenario	Dose	Measurement times	$N_{q'}$	$Q'(Q)$	ϕ_D (E_D ,%)	RSE(%) of (μ, ω^2, σ)		CPU Time (sec)
						VA_0	k	
Reference 26 times	0	0, 7, seq(28, 672, 28)	75	4 (4)	33.7 (100%)	VA_0	1.6, 8.4	FIM 0.06
	150	0, 7, seq(28, 672, 28)	75	k		7.7, 26.4		
	300	0, 7, seq(28, 672, 28)	75	β		12, 16.4		
	500	0, 7, seq(28, 672, 28)	75	E_{max}		16.2, 41.1		
						ED_{50}	47.5	
						σ	0.8	
Fix 26 times optimize doses	0	0, 7, seq(28, 672, 28)	116	3 (7)	38.1 (113.3%)	VA_0	1.6, 8.5	FIM 0.18 optim R 0.50 optim C 0.12
	100	0, 7, seq(28, 672, 28)	50	k		7.2, 22.97		
	500	0, 7, seq(28, 672, 28)	134	β		9.6, 13.1		
				E_{max}		12.99, 35.8		
						ED_{50}	43.8	
						σ	0.8	
Reference 4 times	0	0, 224, 420, 672	75	4 (4)	15.5 (100%)	VA_0	1.6, 9.09	FIM 0.05
	150	0, 224, 420, 672	75	k		14.8, 98.5		
	300	0, 224, 420, 672	75	β		13.6, 19.8		
	500	0, 224, 420, 672	75	E_{max}		17.5, 48.1		
						ED_{50}	50.9	
						σ	3.44	
Fix 4 times optimize doses	0	0, 224, 420, 672	113	3 (7)	17.6 (114.2%)	VA_0	1.6, 9.1	FIM 0.15 optim R 0.25 optim C 0.05
	100	0, 224, 420, 672	56	k		13.6, 84.6		
	500	0, 224, 420, 672	131	β		11.2, 16.1		
				E_{max}		13.9, 41.9		
						ED_{50}	45.4	
						σ	3.47	
Optimize both doses and 4 times	0	0, 168, 196, 672	107	4 (1932)	19.4 (125.5%)	VA_0	1.6, 8.9	FIM 4.96 optim R 431.59 optim C 30.95
	50	0, 7, 644, 672	13	k		11.1, 58.2		
	100	0, 7, 644, 672	48	β		11.5, 16.4		
	500	0, 168, 196, 672	132	E_{max}		13.4, 43.3		
						ED_{50}	44.4	
						σ	3.51	

Table 2: Reference and optimal design for the model of visual acuity, with 26 measurement times on top and 4 measurement times on bottom with 300 subjects. With $N_{q'}$ the number of subjects in each group, Q' the number of elementary designs in the final population design, Q the number of possible elementary designs in the design space, ϕ_D the D-criterion, E_D the D-efficiency, RSE the predicted relative standard error (for fixed effect parameters μ , variance of random effect ω^2 and standard error of the residual error σ) and CPU Time the computing time (FIM calculation running time + optimization running time in R and in C, respectively). Design elements in bold were optimized (for columns "Dose", "Measurement times" and " $N_{q'}$ ").

(among 7 proposed doses) with the 26 same measurement times.

Second, a reference sparse design was evaluated with 4 measurement times quasi-equally spaced between 0 and 672 days and with the same 4 doses groups (0, 150, 300, 500 μg) with the same number of subjects in each group ($\xi_q = (t_q = \{0, 224, 420, 672\}, d_q)$). This design was compared to two optimized designs: one optimizing only the doses among 7 possible doses: 0, 25, 50, 100, 150, 300 and 500 μg and another optimizing both times (with each possible elementary design composed of 2 fixed times: 0 and 672 days and 2 times optimized among 24 proposed times between 7 and 644 days) and doses levels among 7 possible doses: 0, 25, 50, 100, 150, 300 and 500 μg ($\xi_q = (t_q = \{t_1, t_2, t_3, t_4\}, d_q)$).

The multiplicative algorithm parameters were set as follows: $\lambda = 0.99$, $\delta = 0.00001$ and maximum iteration number 10000.

All results were compared in terms of doses, measurement times, number of subjects per group ($N_{q'}$), D-criterion, D-efficiency with respect to the reference design, relative standard errors (%) of all population parameters, number of possible elementary designs in the design space (Q), number of elementary designs in the final population design (Q') and CPU time for running the optimization program. The multiplicative algorithm program can be separated into two parts: first it evaluates FIMs of each elementary design and saves them, then it estimates weights of these elementary designs in regard to their FIMs. CPU time was therefore reported separately for evaluating FIMs and for estimating empirical weights.

4.4. Results

Table 2 reports results of all the designs evaluated and optimized. With a rich design (26 measuring times), optimization of the doses led to a population design composed of 3 groups, which was, as expected, better than the reference design (D-criterion of 38.12 vs 33.66 with the same total number of subjects) and therefore led to a slightly better precision of estimation (for example, the relative standard error (RSE) of ED_{50} decreased from 48% with the non-optimized design to 44% with the optimal design).

Reducing the number of measurements from 26 to 4 in the two reference designs led to a loss of half of the information measured by the D-optimality criterion (from 33.66 to 15.45) whereas the number of measurements was divided by 6.5. With fixed measurement times, the information gain was approximately 14% after optimization of doses. When both doses and measurement times were optimized (last scenario), 1932 elementary designs (or groups) were pre-defined. After optimization of the weights corresponding to these groups, four elementary designs with non-null number of subjects were finally included in the population design. With this design the gain in information was up to 25% as compared to the reference 4 times design (with non-optimized weights). With the optimal design in both dose and times, the estimation precision was also improved for all parameters, and RSE values were below 20% except for ED_{50} (44%). However, the predicted RSE value of ED_{50} was lower than in the non-optimized rich design with 26 samples, whereas the number of visits was 6.5 times lower (with the same total number of subjects).

With respect to CPU time running for the optimization, the more elementary designs were involved (i.e. the larger Q), the longer the function ran. In addition, running time for the evaluation of FIMs was lower than the optimization running time, especially when implemented in R and when Q was large. As seen in Table 2 implementation in C ran much faster than R. The full code of this application can be found in Appendix E.

5. Discussion

In this work, a multiplicative algorithm based on the FIM is proposed and implemented for design optimization for studies with longitudinal data analyzed with nonlinear mixed effect models. The advantages of this algorithm are its simplicity and that it has proof of convergence for several optimality criteria [15]. It increases information after each iteration, which guarantees the quality of its optimal design result, given clinical constraints.

This algorithm was applied to two examples. The first application aims

to design a new study by optimizing both measurement times and doses in a study of a monoclonal antibody indicated in the treatment of wet-AMD. To our knowledge, this is the first time that the multiplicative algorithm is used into such a model with random effects and on top of that, both time measurements and dose administration are optimized. The second application, presented in Appendix A, is an additional test of the algorithm and makes sure that this extension of the algorithm can be applied to any model. This is also the first application of this type of algorithm to a logistic model with repeated discrete data. Here, two specific models and multivariate normal distribution of random effects were considered. However, the multiplicative algorithm could be applied with any structural model and any distribution of random effects as long as the FIM can be computed [27].

The implemented algorithm uses the D-optimality criterion. However, other criteria can be considered. In the A-criterion, the total variance of the parameter estimates is minimized, and this is equivalent to minimizing the average variance [6]. If specific functionals of the parameter, such as the area under the curve, the minimum effective dose or the maximum of the curve, are of interest, the C-criterion is used ($\phi_C = -c^T M c$, where c is a non-zero constant vector) [15, 28]. The D_s -criterion can also be of interest, in order to estimate a subset of the parameters as precisely as possible [6]. They all have proved ability to enable the algorithm to converge to the optimal result monotonically. A straightforward extension of the present work would be to use the algorithm with these different criteria.

Here, we only considered locally optimal designs, which depend on knowledge of the model and parameters. Another possible extension of the multiplicative algorithm is to combine this algorithm with robust criteria, with respect to the model and/or parameters [29, 118].

In conclusion, the extension of the multiplicative algorithm to mixed effect models proposed here allows efficient optimization of population designs, with both continuous and discrete outcomes. The presented version of the algorithm is currently being included in the next release of PFIM.

6. Acknowledgement and declarations

The authors thank Dr. Sebastian Ueckert (Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden) for providing the R code implementing the QRMC-AGQ (quasi-random Monte Carlo - adaptive Gaussian quadrature) method [30].

The authors have no conflict of interest to declare.

References

- [1] P. Van Der Graaf, CPT: Pharmacometrics and Systems Pharmacology, CPT: Pharmacometrics & Systems Pharmacology 1 (9) (2012) e8.
- [2] D. Mould, R. Upton, Basic concepts in population modeling, simulation, and model-based drug development, CPT: Pharmacometrics & Systems Pharmacology 1 (9) (2013) e6.
- [3] D. Mould, R. Upton, Basic concepts in population modeling, simulation, and model-based drug development - part 2: introduction to pharmacodynamic modeling methods, CPT: Pharmacometrics & Systems Pharmacology 2 (4) (2014) e38.
- [4] R. Upton, D. Mould, Basic concepts in population modeling, simulation, and model-based drug development - part 3: introduction to pharmacodynamic modeling methods, CPT: Pharmacometrics & Systems Pharmacology 3 (1) (2014) e88.
- [5] F. Mentré, M. Chenel, E. Comets, J. Grevel, A. Hooker, M. O. Karlsson, M. Lavielle, I. Gueorguieva, Current use and developments needed for optimal design in pharmacometrics: a study performed among DDMoRe's european federation of pharmaceutical industries and associations members, CPT: Pharmacometrics & Systems Pharmacology 2 (2013) e46.

- [6] A. C. Atkinson, A. N. Donev, R. D. Tobias, *Optimum Experimental Designs, with SAS*, Oxford Statistical Science, Oxford University Press 2009, Oxford.
- [7] V. V. Fedorov, S. L. Leonov, *Optimal Design for Nonlinear Response Models*, Chapman and Hall/ CRC Press 2014, Boca Raton.
- [8] L. Pronzato, A. Pázman, *Design of Experiments in Nonlinear Models*, Springer 2013, New York.
- [9] C. Dumont, G. Lestini, H. Le Nagard, F. Mentré, E. Comets, T. T. Nguyen, PFIM Group, PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models, *Computer Methods and Programs in Biomedicine* 156 (2018) 217–229.
- [10] J. Nelder, R. Mead, A Simplex method for function minimization, *Computer Journal* 1 (1965) 308–313.
- [11] V. Fedorov, *Theory of Optimal Experiments*, Academic Press 1972, New York.
- [12] H. P. Wynn, Results in the theory and construction of D-optimum designs, *Journal of the Royal Statistical Society B* 34 (1972) 133–147.
- [13] S. Retout, E. Comets, A. Samson, F. Mentré, Design in nonlinear mixed effects models: optimization using the Federov-Wynn algorithm and power of the Wald test for binary covariates, *Statistics in Medicine* 26 (2007) 5162–5179.
- [14] R. Harman, Computing D-optimal approximate designs of experiments on finite spaces: a survey and comparison of algorithms, Presented at Workshop on Design of Experiments: New Challenges, Marseilles, France (2018).
- [15] Y. Yu, Monotonic convergence of a general algorithm for computing optimal designs, *Annals of Statistics* 38 (3) (2010) 1593–1606.

- [16] P. J. Rosenfeld, D. M. Brown, J. S. Heier, D. S. Boyer, P. K. Kaiser, C. Y. Chung, R. Y. Kim, Ranibizumab for neovascular age-related macular degeneration, *New England Journal of Medicine* 355 (14) (2006) 1419–1431.
- [17] G. Lestini, S. Ueckert, F. Mentré, Robust design in model based analysis of longitudinal clinical data, Presented at PODE meeting, Uppsala, Sweden (2016).
- [18] J. Seurat, T. T. Nguyen, F. Mentré, Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes, *Statistical Methods in Medical Research* 29 (3) (2020) 934–952.
- [19] F. Mentré, A. Mallet, D. Baccar, Optimal design in random effect regression models, *Biometrika* 84 (1997) 429–442.
- [20] T. Mielke, R. Schwabe, Some considerations on the fisher information in nonlinear mixed effects models, *mODa 9—Advances in Model-Oriented Design and Analysis* (2010) 129–136.
- [21] J. Nyberg, C. Bazzoli, K. Ogungbenro, A. Aliev, S. Leonov, S. Duffull, A. Hooker, F. Mentré, Methods and software tools for design evaluation for population pharmacokinetics-pharmacodynamics studies, *British Journal of Clinical Pharmacology* 79 (2014) 6–17.
- [22] T. Holland-Letz, On the combination of c- and D-optimal designs: general approaches and applications in dose-response studies, *Biometrics* 73 (1) (2017) 206–213.
- [23] F. Pukelsheim, *Optimal design of experiments*, SIAM 2006, Philadelphia.
- [24] P. Whittle, Some general points in the theory of optimal experimental design, *Journal of the Royal Statistical Society* (1973).
- [25] A. Pázman, *Foundations of optimum experimental design*, Springer 1986, Dordrecht.

- [26] S. Buatois, S. Ueckert, N. Frey, S. Retout, F. Mentré, Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models, *American Association of Pharmaceutical Scientists Journal* 20 (3) (2018) 56.
- [27] C. Proust, H. Jacqmin-Gadda, Estimation of linear mixed models with a mixture of distribution for the random effects, *Computer Methods and Programs in Biomedicine* 78 (2) (2005) 165–173.
- [28] T. Holland-Letz, H. Dette, D. Renard, Efficient algorithms for optimal designs with correlated observations in pharmacokinetics and dose-finding studies, *Biometrics* 68 (1) (2012) 138–145.
- [29] F. Loingeville, T. T. Nguyen, M.-K. Riviere, F. Mentré, Robust designs in longitudinal studies accounting for parameter and model uncertainties—application to count data, *Journal of Biopharmaceutical Statistics* 30 (1) (2020) 31–45.
- [30] S. Ueckert, F. Mentré, Computation of the Fisher information matrix for discrete nonlinear mixed effect models, Presented at 8th International Conference on Computational and Methodological Statistics, London, UK (2015).
- [31] T. T. Nguyen, F. Mentré, Evaluation of the Fisher information matrix in nonlinear mixed effect models using adaptive Gaussian quadrature, *Computational Statistics & Data Analysis* 80 (2014) 57–69.
- [32] S. Ueckert, F. Mentré, A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature, *Computational Statistics & Data Analysis* 111 (2016) 203–219.

Appendices

Appendix A: Application to design optimization for a longitudinal study with binary outcome

Nonlinear mixed effect models for discrete data

For discrete data, the conditional probability for the vector of observations y_i of individual i is modeled as follows:

$$p(y_i|b_i, z_i, \Psi) = h(y_i, \xi_i, \theta_i) \quad (13)$$

In equation (13), h is the function describing the probability of y_i given individual random effects b_i for a design ξ_i and a vector of subject-specific parameters θ_i . An example of h is given below in this Appendix (with h a logistic model). θ_i is expressed as a function $g(\mu, b_i, z_i, \beta)$ of: the vector of fixed effects μ , the vector of random effects b_i and the vector of covariates z_i for individual i , as well as the vector of covariate effects β . b_i follows a normal distribution with mean zero and variance-covariance matrix Ω which accounts for the inter-individual variability, *i.e.* $b_i \sim \mathcal{N}(0, \Omega)$. Ω_u is the vector containing all unique elements of Ω . The vector Ψ^T contains all population parameters to be estimated ($\Psi^T = (\mu^T, \beta^T, \Omega_u^T)$).

Fisher information matrix by quasi random Monte Carlo-adaptive Gaussian quadrature

When linearization is to be avoided, for example in models with discrete outcomes, alternatives such as adaptive Gaussian quadratures can be used [31, 32]. To evaluate the FIM (equation (4)), one needs to compute two types of integrations: one with respect to b_i and one with respect to y_i . Ueckert and Mentré [32] evaluated the former by AGQ and the latter by MC and showed by CTS that this is a suitable approach with a moderate number of random effects. They then extended this approach by using quasi-random Monte Carlo (QRMC) instead of MC, which leads to huge gains in computing time [30]. The principle of quasi-random MC integration is the following: instead of naively

sampling replicates of b_i from $p(b_i|\Psi)$ then sample replicates of y_i from $p(y_i|b_i)$, samples of b_i in $[0, 1]^p$ were generated from a Sobol sequence and samples were transformed using $p^{-1}(b_i|\Psi)$ and $p^{-1}(y_i, b_i)$.

This approach implemented in an R code is the one used to evaluate the FIM with binary outcomes in this work.

Context and objectives

This example, inspired by a previous work [17, 18], was a one-year clinical trial with repeated binary outcomes (responder vs non-responder) in two balanced treatment groups. The response probability over time was described by a mixed logistic model, including a covariate for the treatment effect. The objective in this application was to determine the optimal location of a limited number of measurement times chosen from 0 to 12 months.

Model and parameters

The model studied describes binary responses ($y = 0$ or 1) recorded over time ($t = 0$ to 12 months) in two different treatment groups. The response probability at time t is given as

$$p(y = 1|b, 1_T) = \frac{\exp(\theta_1 + \theta_2(1 + \beta \times 1_T)t)}{1 + \exp(\theta_1 + \theta_2(1 + \beta \times 1_T)t)}, \quad (14)$$

where θ_1 is the baseline and θ_2 the slope parameter. 1_T is a treatment group indicator variable (with $1_T = 0$ if control group, $1_T = 1$ if experimental group) and β is the treatment effect on θ_2 . θ_1 and θ_2 follow a normal distribution with a diagonal variance matrix: $\theta_l = \mu_l + b_l$ where $b_l \sim \mathcal{N}(0, \omega_l^2)$ for $l = \{1, 2\}$. Figure 3 shows the predicted response curve from 0 to 12 months for each treatment group using nominal values in Table 3.

Design optimization

The multiplicative algorithm was used to optimize the population design Ξ_w , based on the D-optimality criterion. The population FIM was evaluated

Parameter	distribution	μ	ω^2	β
θ_1	normal	-2	0.49	-
θ_2	normal	0.09	0.03	-
β	-	-	-	5

Table 3: Parameter values of the model for repeated binary outcome (with μ the fixed effects, ω^2 the variance of the random effects, and β the treatment effect).

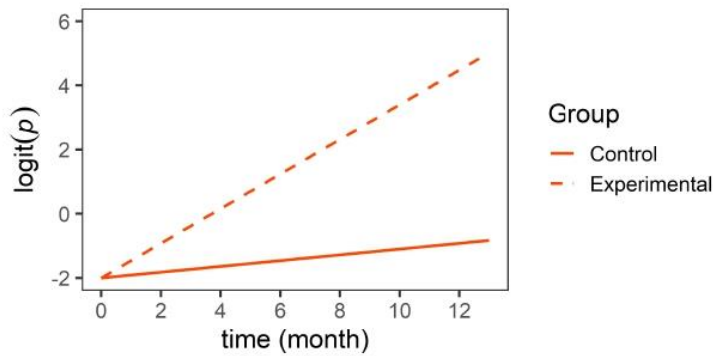


Figure 3: Mean of the logit of the response probability over 12 months predicted by the model (14). The corresponding parameter values are indicated in Table 3. Solid lines represent the control group and dashed lines the experimental group.

using the QRMC/AGQ method. The number of QRMC samples was fixed to 1000 and the number of AGQ nodes to 7.

The population design Ξ was composed of $N = 100$ patients (50 in each treatment group) included in total and assigned to different elementary designs ξ_q , chosen among a finite number Q of possible elementary designs. Measurement times were optimized among several possible times between the beginning (at month 0, fixed) and the end of the study (month 12, fixed). Given these constraints, each elementary design was composed of a vector t_q of four measurement times ($\xi_q = t_q = \{0, t_2, t_3, 12\}$).

First, a reference equispaced non-optimized design was evaluated with 4 measurement times ($\xi_q = (t_q = \{0, 4, 8, 12\})$) and $N = 100$ subjects: 50 in the

two treatment groups.

Second, two out of the four measuring times of the design ($\xi_q = (t_q = \{t_1, t_2, t_3, t_4\})$) were optimized, with the same number of subjects (and same measurement times) in the two treatment groups. Assuming that t_1 and t_4 were fixed to 0 and 12 months (*i.e.* the beginning and the end of the study respectively), only two times t_2 and t_3 were optimized among the set of 11 possible times (1,2,3,4,5,6,7,8,9,10,11). No repetition was considered, *i.e.* four sampling times must be different.

The multiplicative algorithm parameters were set as follows: $\lambda = 0.99$, $\delta = 0.00001$ and maximum iteration number 10000.

The optimal design was thus compared to the reference equispaced design in terms of measurement times, D-efficiency, RSE (notably the treatment effect β , as the expected Wald test power to detect this covariate effect depends on the predicted SE of β) and CPU time. The D-efficiency was calculated in regard to the equispaced design.

Results

Table 4 reports results of the evaluated and optimized designs.

According to the pre-defined clinical constraints, optimal weights of 55 elementary designs were searched by the algorithm. The optimal design was composed of three groups which differed in terms of measurement times. These three groups were respectively composed of 52, 6 and 42 individuals (26, 3 and 21 individuals in each treatment arm, placebo or treated). Compared to the reference design (non-optimized), the information of the optimized design was increased by 18% (as the D-criterion was 34.3 for the equispaced design ($t_2 = 4$ months and $t_3 = 8$ months) and 40.6 for the optimized design), with the same total number of individuals and the same number of measurement times per individual. With the optimal design, predicted RSE was notably improved for the baseline θ_1 and for the treatment effect β parameters.

Compared with the other application (section 4), the CPU time increased, mainly because of the evaluation FIM method (QRMC/AGQ, detailed above).

Scenario	Measurement times	$N_{q'}$	$Q'(Q)$	ϕ_D ($E_D, \%$)	RSE(%) of $(\mu, \omega^2), \beta$		CPU Time (sec)
					θ_1	θ_2	
Reference 4 times	0, 4, 8, 12	100	1 (1)	34.3 (100%)	θ_1 16.1, 156.5 θ_2 35.8, 62.6 β 46.9	FIM 43.4	
Optimize 4 times	0, 2, 3, 12	52	3 (55)	40.6 (118.4%)	θ_1 14.6, 126	FIM 4588.85	
	0, 2, 9, 12	6			θ_2 36.8, 67.6	optim R 4.68	
	0, 3, 7, 12	42			β 42	optim C 0.16	

Table 4: Reference and optimal design for the model for repeated binary outcome, with 4 measurement times and 100 subjects (50 in each treatment group). With $N_{q'}$ the number of subjects, Q' the number of elementary designs in the final population design, Q the number of possible elementary designs in the design space, ϕ_D the D-criterion, E_D the D-efficiency, RSE the predicted relative standard error (for fixed effect parameters μ , variance of random effect ω^2 and treatment effect β) and CPU Time the computing time (FIM calculation running time + optimization running time in R and in C respectively). Design elements in bold were optimized (for columns "Measurement times" and " $N_{q'}$ ").

Once the FIM was evaluated for the 55 designs, the multiplicative algorithm performed in 4.68s by program in R, which reduced to 0.16s in C.

Appendix B: Multiplicative function in R

```

1 multiplicative_algo<-function(FIMs, version, nbParameters,
2   designSpace, nbSubjects, lambda, delta, iteration){
3
4   numberFIMs <- length(FIMs)
5   w<- rep(1/numberFIMs,numberFIMs)
6
7   if( version == "R")
8   {
9     for(k in 1:iteration){
10      #Calculate the sum of matrix with weight associated
11      #idea: Mw<-Mw+w[i]*FIMs[[i]]
12      MFw<-Map(" *",FIMs,w) #associate weights
13      Mw<-Reduce("+",MFw) #sum of all matrices
14      dm<-dim(Mw)[1]
15      Dphi <- det(Mw)^(1/dm) * solve(Mw)/dm # calculate
16      derivatives of function phi_D
17      d<-sapply(FIMs, FUN = function(x) sum(diag(Dphi %*% x)))
18      #the vector of multiplier
19      w<- w*d^lambda/sum(w*d^lambda) #develop weight
20      if(max(d)<(1+delta)*sum(w*d)){break} #stop criterion
21    }
22    Dcriterion <- det(Mw*nbSubjects)^(1/dm)
23  }
24  if( version == "C")
25  {
26    dyn.load("algomult.dll")
27    q<- .C("multiplicativeAlgorithm", originMatrices=as.double(
28      unlist(FIMs)), weights = as.double(w), dim=as.integer(dim(FIMs
29      [[1]])[1]),
30      n=as.integer(numberFIMs), dimA = as.integer(nbParameters
31      ), lambda = as.double(lambda), delta = as.double(delta),
32      it = as.integer(iteration))
33    dyn.unload("algomult.dll")
34    w <- q$weights
35    k <- q$it
36    MFw<-Map(" *",FIMs,w)

```



```

31   Mw<-Reduce("+",MFw)
32   Dcriterion <- det (Mw*nbSubjects)^(1/q$dim)
33 }
34
35 #by default , we take into count weights which give meaningful
   subject numlber
36 v<-which(w>1/nbSubjects) # v<-which(w>mean(w)) ##if we want to
   get weights bigger than the mean
37
38 cat("\n\n*****\n\n")
   n")
39 cat("*** Result of Algorithm *****\n")
40 cat("\n\n IMPORTANT SUBJECT NUMBER *****\n", w[v]*
   nbSubjects)
41 cat("\n\n IMPORTANT DESIGN VALUES *****\n")
42 print (designSpace [, v])
43
44 cat("\n\n NUMBER OF ITERATIONS *****\n",k,
45      "\n\n D-CRITERION *****\n",Dcriterion)
46
47
48 plot (w, ylim=c (0,1))
49 return ( list (w, v))
50 }

```

Appendix C: Multiplicative function in C

```

1 /* FILE NAME: algomult.cpp
2 *
3 * PURPOSE: This file contains function which implement the
4 *           multiplicative algorithm. (Yaming Yu 2010)
5 *
6 * DEVELOPERS: PFIM group, pfim@inserm.fr
7 *             Yuxin Tang, INSERM UMR 1137 BIPID team
8 *
9 * INPUT VARIABLES FOR THE ENTRY POINT FUNCTION (
10 *      multiplicativeAlgorithm) FROM R:
11 *
12 *      Name                Type                Description
13 *      -----                -----                -----
14 *      originMatrices      double*           a list of fisher information
15 *                        matrices(FIM) computed by R functions
16 *                        and read as a vector of values
17 *                        column by column in the following functions
18 *      weights              double*           a vector of weight
19 *                        corresponding to input FIMs and will evolve in the end
20 *                        of each iteration of algorithm
21 *      dim                  int*             the dimension of matrices (all
22 *                        matrices have same dimension)
23 *      n                    int*             total number of input matrices
24 *      dimA                 int*             the dimension of block A of
25 *                        matrix( all matrices have same dimension of
26 *                        block A)
27 *      lambda               double*          a parameter of algorithm
28 *                        between 0 and 1
29 *      delta                double*          a precision of stop criterion
30 *                        of algorithm
31 *      iteration            int*             the algorithm maximum number of
32 *                        iteration
33 *
34 * ALGORITHM
35 * associate each origin FIM input from R with a weight value

```

```

27 * sum all weighted matrices as one matrix
28
29 * calculate determinant of the summed matrix, and also separately
    the determinant of block A and block B
30 of the summed matrix
31
32 * calculate a coefficient as (D-criterion of the summed matrix)
    divided by (the total dimension of a matrix)
33
34 * inverse the summed matrix
35
36 * multiply each element of the inverse matrix with the
    coefficient, the matrix obtained is the derivative
37 of phi
38
39 * multiply every origin FIM input from R with the derivative of
    phi, then compute the trace of each matrix
40 product
41
42 * evolve weights by multiply corresponding (trace^lambda), and
    divided by its sum
43
44 * algorithm stops when criterion about delta is satisfied
45 */
46
47
48 #include <utility>
49 #include <limits.h>
50 #include <math.h>
51 #include "algomult.h"
52
53
54 #ifdef _WINDOWS
55
56 #include <Windows.h>
57
58 // windows dll entry point
59 BOOL WINAPI DllMain( HMODULE hModule, DWORD ul_reason_for_call,

```

```

        LPVOID lpReserved )
60 {
61     switch (ul_reason_for_call)
62     {
63     case DLL_PROCESS_ATTACH:
64     case DLL_THREAD_ATTACH:
65     case DLL_THREAD_DETACH:
66     case DLL_PROCESS_DETACH:
67         break;
68     }
69     return TRUE;
70 }
71
72 #elif
73
74
75 #endif
76
77 /* FUNCTION NAME: sum
78 *
79 * DESCRIPTION: This function adds up a list of matrices as one
80 *              matrix.
81 *
82 * INPUT:
83 * Name          Type          Description
84 * -----
85 * matrices      double*       a vector of n matrices column
86 *               by column of dimension dim
87 * size          int           total number of element of a
88 *               matrix (size = dim * dim)
89 * n             int           total number of matrices
90 *
91 * RETURN VALUE:
92 * Name          Type          Description
93 * -----
94 * result        double*       one matrix where each element
95 *               equals the sum of element of input matrices at
96 *               that position

```

```

93  */
94  void sum(double *matrices , int size , int n, double *result)
95  {
96      int i, j;
97      // initialise the result matrix
98      // copy the first matrix in the list to the result matrix
99      for (j = 0; j < size; j++) {
100         *result = *matrices;
101         result++;
102         matrices++;
103     }
104     result -= size;
105
106     // add rest matrices in the list to the result matrix
107     for (i = 1; i < n; i++) {
108         for (j = 0; j < size; j++){
109             *result += *matrices;
110             result++;
111             matrices++;
112         }
113         result -= size;
114     }
115 }
116
117 /* FUNCTION NAME:  MatrixTimesWeight
118  *
119  * DESCRIPTION: This function multiplies each element in the matrix
120                 by a weight.
121  *
122  * INPUT:
123  * -----
124  * matrices      double*      a vector of n matrices column
125                 by column of dimension dim
126  * weights       double*      n values, one for each matrix
127  * size          int          total number of element of a
128                 matrix (size = dim * dim)
129  * n             int          total number of matrices

```

```

128 *
129 * RETURN VALUE:
130 * Name          Type          Description
131 * -----
132 * result        double*       weighted matrices, same number
133 *               of matrices as input
134 */
135 void MatrixTimesWeight(double *matrices, double *weights, int size,
136                       int n, double *result)
137 {
138     int size_n = size * n;
139     for (int i = 0; i < size_n; i++)
140     {
141         *result = (*matrices) * (*weights);
142         matrices++;
143         result++;
144         if ((i + 1) % size == 0)
145             weights++;
146     }
147 }
148
149
150 /* FUNCTION NAME:  decompositionLU
151 *
152 * DESCRIPTION: This function decompose a matrix as a lower
153 *              triangular matrix and an upper triangular matrix.
154 *
155 * INPUT:
156 * Name          Type          Description
157 * -----
158 * matrix        double*       a symmetric matrix, can be
159 *               divided into two blocks independent non-zero
160 *               (block A,B)
161 * dim           int           dimension of an entire matrix
162 * dimA          int           dimension of block A of the
163 *               matrix (dimension of block B = dim - dimA)

```

```

161 *
162 * RETURN VALUE:
163 * Name          Type          Description
164 * -----
165 * L              double*       LU decomposition result —
    lower matrix
166 * U              double*       LU decomposition result —
    upper matrix
167 */
168 void decompositionLU(double *matrix, double *L, double *U, int dim,
    int dimA) {
169     int i = 0, j, k, i_dim, i_j_dim, jj_dim;
170
171     // decomposition LU for block A
172     do {
173         i_dim = i * dim;
174         for (j = 0; j < i; j++) {
175             jj_dim = j * dim + j;
176             if (U[jj_dim] == 0) {
177                 i = dimA;
178                 break;
179             }
180             i_j_dim = i_dim + j;
181             L[i_j_dim] = matrix[i_j_dim];
182             for (k = 0; k < j; k++) {
183                 L[i_j_dim] -= L[i_dim + k] * U[k * dim + j];
184             }
185             L[i_j_dim] /= U[jj_dim];
186
187         }
188         for (j = i; j < dimA; j++) {
189             i_j_dim = i_dim + j;
190             U[i_j_dim] = matrix[i_j_dim];
191             for (k = 0; k < i; k++) {
192                 U[i_j_dim] -= L[i_dim + k] * U[k * dim + j];
193             }
194         }
195         i++;

```

```

196 } while (i < dimA);
197
198 // decomposition LU for block B, if dimA < dim
199 while (i < dim) {
200     i_dim = i * dim;
201     for (j = dimA; j < i; j++) {
202         jj_dim = j * dim + j;
203         if (U[jj_dim] == 0) {
204             i = dim;
205             break;
206         }
207         i_j_dim = i_dim + j;
208         L[i_j_dim] = matrix[i_j_dim];
209         for (k = dimA; k < j; k++) {
210             L[i_j_dim] -= L[i_dim + k] * U[k * dim + j];
211         }
212         L[i_j_dim] /= U[jj_dim];
213     }
214     for (j = i; j < dim; j++) {
215         i_j_dim = i_dim + j;
216         U[i_j_dim] = matrix[i_j_dim];
217         for (k = dimA; k < i; k++) {
218             U[i_j_dim] -= L[i_dim + k] * U[k * dim + j];
219         }
220     }
221     i++;
222 };
223
224
225
226 }
227
228
229 /* FUNCTION NAME:  inverse
230 *
231 * DESCRIPTION: This function inverts a symmetric matrix
232 *
233 * INPUT:

```



```

234 * Name          Type          Description
235 * -----
236 * m             double*       a symmetric matrix to be
                inverted, can be divided into two independent
237 *                                     blocks (A, B)
238 * dim           int           dimension of an entire matrix
239 * dimA          int           dimension of block A of the
                matrix (dimension of block B = dim - dimA)
240 * detA          double        determinant of block A
241 * detB          double        determinant of block B
242 *
243 * RETURN VALUE:
244 * Name          Type          Description
245 * -----
246 * invM          double*       inversed matrix
247 */
248 void inverse(double *m, int dim, double *invM, int dimA, double
                detA, double detB)
249 {
250     int i, j, ii, jj, il, jl, diml, il_dim, ii_dim, k;
251     double det, *c, *d, *l, *u;
252
253     // prepare matrix L and U
254     l = (double *)calloc(dim * dim, sizeof(double));
255     u = (double *)calloc(dim * dim, sizeof(double));
256     // initialise matrix L with all diagonal
257     for (i = 0; i < dim; i++)
258         l[i*dim + i] = 1;
259
260     //calculate inverse for block A
261     diml = dimA - 1;
262     c = (double *)calloc((diml) *(diml), sizeof(double));
263
264     for (j = 0; j < dimA; j++) {
265         for (i = 0; i < dimA; i++) {
266             il = 0;
267             for (ii = 0; ii < dimA; ii++) {
268                 if (ii == i)

```

```

269     continue;
270     il_dim = il * dim1;
271     ii_dim = ii * dim;
272     j1 = 0;
273     for (jj = 0; jj < dimA; jj++) {
274         if (jj == j)
275             continue;
276         *(c + il_dim + j1) = *(m + ii_dim + jj);
277         j1++;
278     }
279     il++;
280 }
281
282 decompositionLU(c, l, u, dim1, dim1);
283 det = 1;
284 for (k = 0; k < dim1; k++)
285     det *= u[k * dim1 + k];
286 invM[i + j * dim] = pow(-1.0, i + j + 2.0) * (det) / (detA);
287 }
288 }
289
290 //calculate inverse for block B
291 dim1 = dim - dimA - 1;
292 d = (double *)calloc((dim1) *(dim1), sizeof(double));
293
294 for (j = dimA; j < dim; j++) {
295
296     for (i = dimA; i < dim; i++) {
297         il = 0;
298         for (ii = dimA; ii < dim; ii++) {
299             if (ii == i)
300                 continue;
301             il_dim = il * dim1;
302             ii_dim = ii * dim;
303             j1 = 0;
304             for (jj = dimA; jj < dim; jj++) {
305                 if (jj == j)
306                     continue;

```

```

307     *(d + i1_dim + j1) = *(m + ii_dim + jj);
308     j1++;
309 }
310 i1++;
311 }
312
313 decompositionLU(d, l, u, dim1, dim1);
314 det = 1;
315 for (k = 0; k < dim1; k++)
316     det *= u[k * dim1 + k];
317 invM[i + j * dim] = pow(-1.0, i + j + 2.0) * (det) / (detB);
318
319 }
320 }
321 free(c);
322 free(d);
323 free(l);
324 free(u);
325 }
326
327
328 /* FUNCTION NAME: computeTrace
329 *
330 * DESCRIPTION: This function multiplies each matrix in the list by
331 *               the same matrix, then compute each
332 *               product's trace
333 *
334 * INPUT:
335 * Name          Type          Description
336 * -----
337 * matrices      double*       a list of matrices to produce
338 *               with theMatrix
339 * theMatrix     double*       the matrix to be multiplied
340 *               with each matrices in the list
341 * dim           int           dimension of an entire matrix
342 * n             int           number of matrices in the list
343 *
344 * RETURN VALUE:

```

342	* Name	Type	Description
343	*		
344	* trace	double*	the trace vector of each matrix product
345	*/		
346	void computeTrace(double *matrices, double *theMatrix, int dim, int n, double *trace)		
347	{		
348	int im, il, in, i;		
349	double *productMatrices = NULL;		
350	int size = dim * dim, i_size;		
351	// intermediate matrices		
352	productMatrices = (double *)calloc(n * size, sizeof(double));		
353			
354	for (i = 0; i < n; i++) {		
355	*trace = 0;		
356	for (in = 0; in < dim; in++)		
357	{		
358	for (im = 0; im < dim; im++)		
359	{		
360	*productMatrices = 0;		
361	for (il = 0; il < dim; il++)		
362	{		
363	*productMatrices += (*matrices) * (*theMatrix);		
364	matrices++;		
365	theMatrix += dim;		
366	}		
367	matrices -= dim;		
368	theMatrix -= size;		
369	if (im == in)		
370	*trace += *productMatrices;		
371	productMatrices++;		
372	theMatrix++;		
373	}		
374	matrices += dim;		
375	theMatrix -= dim;		
376	}		
377	trace++;		

```

378 }
379 productMatrices == n * size;
380 free (productMatrices);
381 }
382
383
384 /* FUNCTION NAME:  multiplicativeAlgorithm
385 *
386 * DESCRIPTION: This is the main function of multiplicative
387 *              algorithm ,which is exported and used in PFIM
388 *              implemented in R-S4
389 *
390 * INPUT:
391 * Name          Type          Description
392 * -----
393 * originMatrices  double*      a list of matrices computed by
394 *                   R functions
395 * weights         double*      a vector of weight
396 *                   corresponding to each matrix in the originMatrices list
397 * dim             int          dimension of an entire matrix (
398 *                   all same for matrices in the list)
399 * n               int*         number of matrices in the list
400 * dimA            int*         dimension of block A
401 * lambda          double*      convergence power of the
402 *                   algorithm (0, 1]
403 * delta           double*      algorithm stop criterion
404 *                   parameter
405 * iteration       int*         algorithm maximum iteration
406 *
407 * RETURN VALUE:
408 * Name          Type          Description
409 * -----
410 * weights       double*      the vector of weights evolved
411 *                   by algorithm
412 * iteration     int*         real iteration run number
413 */
414 void multiplicativeAlgorithm(double *originMatrices , double *
415 weights , int *dim,

```

```

408 int *n, int *dimA, double *lambda, double *delta, int *iteration
    )
409 {
410 double det, coef, sunw, sumwd, maxTrace, *weightedMatrices, *
    sumWeightedMatrices,
411 *inverseSumWeightMat, *Dphi, *tr, *l, *u, detA, detB;
412 int idim = *dim, in = *n, idimA = *dimA, isize = idim * idim, i,
    it;
413
414 weightedMatrices = (double *)calloc(in * isize, sizeof(double));
415 if (!weightedMatrices) {
416     it = -2;
417     return;
418 }
419 sumWeightedMatrices = (double *)calloc(isize, sizeof(double));
420 inverseSumWeightMat = (double *)calloc(isize, sizeof(double));
421 Dphi = (double *)calloc(isize, sizeof(double));
422 tr = (double *)calloc(in, sizeof(double));
423 l = (double *)calloc(isize, sizeof(double));
424 u = (double *)calloc(isize, sizeof(double));
425
426 // initialise diagonal elements of the low matrix for
    decomposition LU as 1
427 for (i = 0; i < idim; i++)
428     l[i*idim + i] = 1;
429
430 for (it = 0; it < (*iteration); it++)
431 {
432     // associate each FIM calculated in R with a weight value
433     MatrixTimesWeight(originMatrices, weights, isize, in,
        weightedMatrices);
434     // sum all weighted matrices
435     sum(weightedMatrices, isize, in, sumWeightedMatrices);
436
437     //calculate determinant of Mw, and separately of block A and B
        of Mw
438     detA = 1;
439     detB = 1;

```

```

440 decompositionLU(sumWeightedMatrices, l, u, idim, idimA);
441 for (i = 0; i < idimA; i++)
442     detA *= u[i * idim + i];
443 for (i = idimA; i < idim; i++)
444     detB *= u[i * idim + i];
445 det = detA * detB;
446 if (det < 0 || isnan(det)) {
447     it = -1;
448     break;
449 }
450
451 // calculate a coefficient as (D-criterion)/(dimension of
452 // matrix)
453 coef = pow(det, (1.0 / idim)) / idim;
454 // inverse the sum matrix of weighted matrices
455 inverse(sumWeightedMatrices, idim, inverseSumWeightMat, idimA,
456 detA, detB);
457 // multiply each element of the inverse matrix with the
458 // coefficient,
459 // the result obtained is the derivative of phi
460 MatrixTimesWeight(inverseSumWeightMat, &coef, isize, 1, Dphi);
461 // multiply every FIM calculated in R with the derivative of
462 // phi,
463 // then obtain the trace of each product
464 computeTrace(originMatrices, Dphi, idim, in, tr);
465 // evolve weights
466 sumw = 0.0; // sum of weights
467 sumwd = 0.0; // sum of lambda exponential weights
468 for (i = 0; i < in; i++)
469 {
470     sumwd += weights[i] * tr[i];
471     weights[i] *= pow(tr[i], *lambda);
472     sumw += weights[i];
473 }
474 maxTrace = tr[0];
475 for (i = 0; i < in; i++)
476 {
477     weights[i] /= sumw; // make sure weights are proper

```

```
fractions
474     if (maxTrace < tr[i])
475         maxTrace = tr[i];
476     }
477
478     if (maxTrace < (1 + *delta)*sumwd)    // stop criterion of the
algorithm
479     {
480         printf("loop finished with iteration = %d\n", it);
481         break;
482     }
483
484 }
485 *iteration = it;
486
487
488 free (weightedMatrices);
489 free (sumWeightedMatrices);
490 free (inverseSumWeightMat);
491 free (Dphi);
492 free (tr);
493 free (l);
494 free (u);
495 }
```


Appendix D: Fisher Information Matrix evaluation function

```

1 funFIMpop <- function(equation , parameters , beta , o , sigma , t_group ,
  Trand , d , PropSubjects , nbTot){
2   #List of names of the fixed effects parameters and sampling times
3   ParametersInEquation<-c(parameters , "t")
4
5   PSI<-c(parameters , "sig.inter" , "sig.slope")
6   lpsi<-length(PSI)
7   lengthParameters<-length(parameters)
8
9   #(Diagonal Matrix of) variance for inter-subject random effects:
10  omega<-diag(o)
11
12  #Parameters of standard deviation model of residual error (sig.
    inter+sig.slope*f)^2:
13  sig.inter<-sigma[1]
14  sig.slope<-sigma[2]
15
16  #get derivatives of sigma
17  Vmodel<- parse(text = paste0("( sig.inter + sig.slope * fixed )
    ^2"))
18  dv<-deriv(Vmodel[[1]] , PSI)
19
20  #gather all groups of protocol
21  M_f<-list()
22  M.F <- matrix(rep(0) , nrow=lpsi+lengthParameters , ncol=lpsi+
    lengthParameters)
23  for(q in 1:length(t_group)){
24    #design of sampling times
25    t<-t_group[[q]]
26
27    #dose value
28    dose<-d[q]
29
30    #calculate matrix E for n individuals
31    equatf <- parse(text = equation , n=-1)
32    f<-function(ParametersInEquation){eval(equatf[[1]])}

```

```

33
34 #Fixed effects parameters values
35 for(i in 1:lengthParameters){
36   assign(parameters[i],beta[i])
37 }
38
39 #calculate the observations with personal parameters
40 parameterValues <- c(beta,t)
41 fixed<-f(parameterValues)
42
43 #calculate variance Var
44 var<-diag((sig.inter+sig.slope*fixed)^2)
45
46 #get derivatives for fixed parameters
47 df<-deriv(equatf[[1]],PSI)
48 mdf<-attributes(eval(df))$gradient
49 #delete the last two columns (correspond to sig.inter and sig.
50 slope)
51 mdfi <- mdf[,-c(lpsi-1,lpsi)]
52 #complete derivative for random effect model
53 #Random effect model (1) = additive (2) = exponential
54
55 #
56
57
58 beta0 <- beta
59 beta0[which(Trand==1)] <- 1
60 mdfie <- mdfi %*% diag(beta0)
61
62 #calculate variance Vi
63 Vi <- mdfie %*% omega %*% t(mdfie) + var
64 #inverse of matrix Vi
65 SVi <- solve(Vi)
66 #compute derivatives of sigma
67 mdv<-attributes(eval(dv))$gradient
68
69
70 #calculate matrix part A
71 MA <- t(mdfi) %*% SVi %*% mdfi

```

```

68
69 #complete the rest of the matrix with 0
70 MA <- cbind(MA, matrix(rep(0), ncol=lpsi, nrow=lengthParameters)
71 )
72 MA <- rbind(MA, matrix(rep(0), ncol=lpsi+lengthParameters, nrow=
73 lpsi))
74
75 #calculate matrix part B
76 #initialize the matrix with 0
77 MB <- matrix(rep(0), nrow=lpsi+lengthParameters, ncol=lpsi+
78 lengthParameters)
79 #prepare a list of matrix of derivatives of sigma to simplify
80 usage
81 if (length(t)==1){
82   m <- lapply(c((lpsi-1), lpsi), function(i, mdv) mdv[i], mdv=mdv )
83 } else {
84   m <- lapply(c((lpsi-1), lpsi), function(i, mdv) diag(mdv[, i]),
85 mdv=mdv )
86 }
87
88 #calculate first three rows of part B
89 for(i in 1:lengthParameters){
90   mdfie <- (mdfie[, i] %*% t(mdfie[, i]))
91
92   MB[lengthParameters+i, seq(lengthParameters+1,
93 lengthParameters+lengthParameters)] <- sapply(
94     lapply(seq(1, lengthParameters), function(i, mdfie) mdfie[, i]
95 ), mdfie=mdfie),
96     function(x) 1/2 * sum(diag((mdfie) %*% SVi %*% (x %*% t(x)
97 ) %*% SVi)))
98
99   MB[lengthParameters+i, c(lengthParameters+lpsi-1,
100 lengthParameters+lpsi)] <- sapply(
101     m, function(x) 1/2 * sum(diag((mdfie) %*% SVi %*% x %*% SVi
102 )))
103 }
104 #calculate the last two rows of partB
105 for(i in (lpsi-1):lpsi){

```

```

96     MB[lengthParameters+i, seq(lengthParameters+1,
97     lengthParameters+lengthParameters)] <- sapply(
98     lapply(seq(1,lengthParameters), function(i,mdfie) mdfie[,i
99     ],mdfie=mdfie),
100     function(x) 1/2 * sum(diag((m[[i-lengthParameters]]) %*%
101     SVi %*% (x %*% t(x)) %*% SVi)))
102
103     MB[lengthParameters+i, c(lengthParameters+lpsi-1,
104     lengthParameters+lpsi)] <- sapply(
105     m, function(x) 1/2 * sum(diag(((m[[i-lengthParameters]]))
106     %*% SVi %*% x %*% SVi)))
107     }
108
109     M.f[[q]] <- (MA+MB)*PropSubjects[q]
110     M.F <-M.F+M.f[[q]]
111   }
112   M.F <- M.F *nbTot
113
114   #set names for vectors
115   fname <-c(sapply(1:lengthParameters, function(x) paste0("u_",
116   parameters[x])),sapply(
117   1:lengthParameters, function(x) paste0("w2_",parameters[x])),
118   "sig.inter", "sig.slope")
119   rownames(M.F) <- fname
120   colnames(M.F) <- fname
121
122   if(0 %in% c(o, sigma)){
123     M.F<-M.F[-c(lengthParameters+which(c(o, sigma)==0)),-c(
124     lengthParameters+which(c(o, sigma)==0))]
125     PSI<- PSI[-c(lengthParameters+which(c(o, sigma)==0))]
126   }
127
128   if(length(t)==1 ){
129     return(list(M.F, det(M.F)))
130   }else {
131     tryCatch({
132       deterFim <- det(M.F)

```

```

126 SE <- sqrt(diag(solve(MF)))
127 RSE <- 100 * SE / c(beta,o,sigma)[which(c(beta,o,sigma)!=0)]
128 CritereDopt <- deterFim^(1/length(c(beta,o,sigma)[which(c(
beta,o,sigma)!=0)]))
129 #write the output in console
130 #      cat("***** FISHER INFORMATION MATRIX
*****\n")
131 #      print(MF)
132 #
133 #      cat("\n\n***** DETERMINANT OF THE
MATRIX *****\n", deterFim,
134 #          "\n\n***** CRITERION
*****\n", CritereDopt,
135 #          "\n\n***** STANDARD ERROR
*****\n", SE,
136 #          "\n\n***** RELATIVE STANDARD ERROR
*****\n", RSE, "\n\n")
137 return(list(MF, deterFim, CritereDopt, SE, RSE))
138 }, error=function(e){
139 return(list(MF, det(MF)))
140 })
141
142 # draw the graphical illustration of model
143 # t<-seq(min(unlist(t_group)),max(unlist(t_group)),(max(
unlist(t_group))-min(unlist(t_group)))/100)
144 # x<-c(beta,t)
145 # fvalue<-f(x)
146 # lines(t,fvalue) #then plot the line of function
147 }
148
149 }

```

Appendix E: Application of multiplicative algorithm

```

1 model <- "exp(VA)+(1-exp(-exp(k)*t))*(emax*dose/(exp(ED50)+dose)-
  exp(be)*exp(VA))"
2 paramName <- c("VA","k","be","emax","ED50")
3 mu <- c(log(55),log(0.005),log(0.2),30,log(150))
4 omega <- c(0.07,0.5,1,150,0)
5 sigma <- c(sqrt(28),0)
6 allTimes <- c(0,7,seq(28,672,28))
7 equiSpacedTimes <- c(0,224,420,672)
8 fixed_times <- list(c(1,0),c(4,672))
9 Trand <- c(1,1,1,1,1)
10 dose <- c(0,25,50,100,150,300,500)
11 lambda <- 0.99
12 delta <- 0.00001
13 iteration <- 10000
14 nbSubjects <- 300
15
16 #####
17 # Reference 26 times
18 start <- Sys.time()
19 ResultEval <- funFIMpop(model, paramName, mu, omega, sigma, list(
  allTimes, allTimes, allTimes, allTimes),
20 Trand, c(0,150,300,500),c(75,75,75,75)/
  nbSubjects, nbSubjects)
21 end <- Sys.time()
22 end-start
23
24 # Fix 26 times and optimize doses
25 # Create design space and elementary FIM corresponding
26 start <- Sys.time()
27 nbTimes <- 26
28 Designs_ini <- combn(allTimes, nbTimes)
29 FIMs <- list()
30 Designs <- matrix(nrow=nbTimes+1)
31 n <- 1
32 for(i in 1:dim(Designs_ini)[2]){

```

```

33   for(j in 1:length(dose)){
34     #the number of design groups need to change in accordance
with the number of doses
35     FIMs[[n]] <-funFIMpop(model,paramName,mu,omega,sigma,list(
Designs_ini[i]),Trand,dose[j],1,1)[[1]]
36     Designs<-cbind(Designs,c(dose[j], Designs_ini[i]))
37     n<-n+1
38   }
39 }
40 Designs <- Designs[,-1]
41 end <- Sys.time()
42 end-start
43
44 # Optimize using multiplicative algorithm implemented in R or C
45 start <- Sys.time()
46 weights <- multiplicative_algo(FIMs, "R", length(paramName),
Designs, nbSubjects, lambda, delta, iteration)
47 end <- Sys.time()
48 end-start
49
50 #####
51 # Reference 4 equi-spaced times
52 start <- Sys.time()
53 ResultEval <- funFIMpop(model, paramName, mu, omega, sigma, list(
equiSpacedTimes, equiSpacedTimes,
54
equiSpacedTimes, equiSpacedTimes),
55
Trand, c(0,150,300,500),c(75,75,75,75)/
nbSubjects, nbSubjects)
56 end <- Sys.time()
57 end-start
58
59 # Fix 4 equi-spaced times and optimize doses
60 # Create design space and elementary FIM corresponding
61 start <- Sys.time()
62 nbTimes <- 4
63 Designs_ini<-combn(equiSpacedTimes, nbTimes)

```

```

64 FIMs<-list ()
65 Designs<-matrix (nrow=nbTimes+1)
66 n<-1
67 for(i in 1:dim(Designs_ini)[2]){
68   for(j in 1:length(dose)){
69     #the number of design groups need to change in accordance
with the number of doses
70     FIMs [[n]] <-funFIMpop(model ,paramName ,mu, omega , sigma , list (
Designs_ini [ , i ]),Trand,dose [j],1,1) [[1]]
71     Designs<-cbind(Designs ,c(dose [j] , Designs_ini [ , i ]))
72     n<-n+1
73   }
74 }
75 Designs <- Designs[,-1]
76 end <- Sys.time ()
77 end-start
78
79 # Optimize using multiplicative algorithm implemented in R or C
80 start <- Sys.time ()
81 weights <- multiplicative_algo(FIMs, "R", length(paramName),
Designs , nbSubjects , lambda , delta , iteration)
82 end <- Sys.time ()
83 end-start
84
85 #####
86 # Optimize both doses and 4 times
87 # Create design space and elementary FIM corresponding
88 start <- Sys.time ()
89 nbTimes <- 4
90 Designs_ini<-combn(allTimes ,nbTimes)
91 if(length(fixed_times)>0){
92   for(f in 1:length(fixed_times)){
93     Designs_ini<-Designs_ini[ ,which(Designs_ini [ fixed_times [[ f
]][1] ,)=fixed_times [[ f ]][2] )
94   }
95 }
96

```



```

97 FIMs<-list ()
98 Designs<-matrix (nrow=nbTimes+1)
99 o<-1
100 for(l in 1:dim(Designs_ini)[2]){
101   for(m in 1:length(dose)){
102     #the number of design groups need to change in accordance
with the number of doses
103     FIMs[[o]] <-funFIMpop(model,paramName,mu,omega,sigma,list(
Designs_ini[,l]),Trand,dose[m],1,1)[[1]]
104     Designs<-cbind(Designs,c(dose[m],Designs_ini[,l]))
105     o<-o+1
106   }
107 }
108 Designs <- Designs[,-1]
109 end <- Sys.time()
110 end-start
111
112 # Optimize using multiplicative algorithm implemented in R or C
113 start <- Sys.time()
114 weights <- multiplicative_algo(FIMs, "R", length(paramName),
Designs, nbSubjects, lambda, delta, iteration)
115 end <- Sys.time()
116 end-start

```

2.2.2. Contribution à la nouvelle version de PFIM

Le programme PFIM (www.pfim.biostat.fr) est conçu pour l'évaluation et l'optimisation de protocoles d'études longitudinales basées sur la FIM (individuelle, de population ou bayésienne). Sa version actuellement disponible en ligne est la version 4.0. (Dumont et al. 2018). Cette version, codée en R par plusieurs membres du groupe PFIM successifs, inclut beaucoup de fonctionnalités mais sa conception peut entraîner des difficultés à ajouter de nouvelles méthodes, voire à utiliser PFIM pour des modèles ou des protocoles complexes.

Une nouvelle version de PFIM, PFIM 5.0., ré-écrite entièrement, est donc en cours de développement. L'objectif de ce travail est d'obtenir un programme modulable, de manière à faciliter l'évolutivité et l'utilisation de PFIM. Pour cela, PFIM est écrit en R S4, langage orienté-objet. La programmation orientée-objet (POO) offre certains avantages par rapport à la programmation traditionnelle (procédurale). En concevant le programme en ensemble d'objets qui communiquent entre eux, la POO permet notamment de gagner en compréhensibilité, d'éviter les duplications de code et de faciliter la réutilisation.

PFIM 5.0. est donc organisé en différentes classes et différents objets. Comme PFIM 4.0., il permet l'évaluation ou l'optimisation de protocoles individuels, de population et bayésiens pour des modèles à une ou plusieurs réponses, en solution analytique ou sous forme d'équations différentielles ordinaires, provenant de la librairie de modèles ou définis par l'utilisateur. Certaines fonctionnalités qui n'étaient pas incluses dans PFIM 4.0. sont disponibles. Il est désormais possible d'utiliser des distributions des effets aléatoires différentes (normale et log-normale) au sein du même projet. L'algorithme multiplicatif (section 2.2.1.) est intégré à PFIM 5.0. en langage C (plus performant que R, ce qui facilitera la résolution des problèmes d'optimisation les plus complexes). L'utilisation de modèles PK de perfusion et des administrations multiples a été simplifiée par rapport à PFIM 4.0. Les graphiques présentés ont aussi été revus, avec utilisation de la librairie R *ggplot2*. Certaines fonctionnalités présentes dans PFIM 4.0. doivent encore être ajoutées, par exemple, l'optimisation de protocoles par l'algorithme de Fedorov-Wynn et l'inclusion des effets de covariables catégorielles. La nouvelle version de PFIM ainsi développée sera prochainement disponible sur le CRAN (*Comprehensive R Archive Network*). Des mises à jour futures pourront aussi permettre d'inclure les récents développements en optimisation de protocoles tels que l'évaluation de la FIM par des méthodes adaptées aux modèles de données discrètes ou encore des méthodes de planification tenant compte de l'incertitude des paramètres et du modèle (section 2.1.).

En tant que membre actif du groupe PFIM, mes rôles incluent le traitement des requêtes d'utilisateurs de PFIM 4.0. (version R et interface). Concernant l'élaboration de PFIM 5.0., mes activités principales sont la participation aux réunions hebdomadaires et l'élaboration d'exemples dans le but de vérifier l'ajout de nouvelles méthodes au logiciel ou pour être intégrés au futur guide d'utilisateur. Un poster à propos de la nouvelle version de PFIM a aussi été produit (voir Annexes) et notamment présenté au PAGE (*Population Approach Group of Europe*), un congrès de pharmacométrie (Seurat et al. 2019).

Une version alpha de PFIM 5.0. est actuellement disponible en interne, une version beta sera disponible pour un nombre plus large d'utilisateurs durant le premier semestre de l'année 2021, dans le but de soumettre une version finale de PFIM 5.0. au cours du deuxième semestre de l'année 2021.

2.3. Comparaison de protocoles pour évaluer la décroissance tumorale précoce

2.3.1. Résumé

Le choix de protocole est complexe dans les études cliniques en oncologie, particulièrement dans les phases cliniques précoces du développement des anticancéreux. Dans ces études, il est de plus en plus recommandé de comprendre la relation entre la concentration des médicaments et la réponse clinique (Smoragiewicz et al. 2020). Cette dernière peut être suivie par des biomarqueurs tels que la décroissance tumorale précoce, c'est-à-dire l'effet des premières semaines de thérapie sur la taille tumorale. Nous avons donc évalué et comparé différents protocoles à partir de simulations, dans le cadre de l'étude d'une combinaison de deux anticorps monoclonaux (associé au traitement standard, la chimiothérapie) pour le traitement du cancer colorectal en phase précoce clinique.

Des jeux de données de tailles tumorales (la taille tumorale est en fait la somme des plus longs diamètres de chaque lésion) ont été simulés à partir de modèles d'inhibition de la croissance tumorale, proches du modèle de Claret (Claret et al. 2009). Trois différents scénarios ont été étudiés concernant l'interaction entre les deux médicaments étudiés : le cetuximab (C, traitement de référence) et une nouvelle molécule (M). Dans le premier scénario, nous avons considéré un effet nul de M dans le but d'évaluer l'erreur de type 1. Dans les deux autres scénarios, nous avons considéré respectivement une simple additivité ou une synergie d'effet entre C et M. Pour chacun de ces scénarios, trois protocoles d'études ont été évalués : à un bras,

deux bras et quatre bras de traitement, avec un même nombre total d'individus ($N=60$). Pour chaque protocole et chaque scénario, 500 jeux de données ont été générés et des mesures de taille tumorale à l'initiation de traitement (semaine 0) puis toutes les deux semaines jusqu'à la semaine 8. Ces jeux de données ont ensuite été analysés par modèles non-linéaires à effets mixtes, avec une estimation des paramètres par l'algorithme SAEM de Monolix 2018R2. Le critère d'évaluation de la combinaison était fondé sur la décroissance de taille tumorale à 8 semaines (ETS8), c'est-à-dire l'écart relatif entre la taille tumorale à la semaine 8 par rapport à celle de la semaine 0. Les tests statistiques, présentés dans le paragraphe suivant, ont été réalisés pour comparer, entre les bras, les distributions d'ETS8, soit observées, soit prédites à partir des *Empirical Bayes estimates* (EBE) obtenus par modélisation.

Pour le design à un bras, les 60 individus ont été traités par la combinaison C+M aux doses standard, les ETS8 de ces individus ont été comparées par test de Wilcoxon à une valeur de référence obtenue lors d'une étude antérieure. Pour le design à deux bras, 30 individus ont été traités par la combinaison C+M aux doses standard et 30 par le traitement de référence (C). Les ETS8 ont été comparés entre les deux groupes par test de Wilcoxon. Pour le design à quatre bras (15 individus par bras), en plus des deux groupes C et C+M aux doses standards, des groupes de combinaisons C+M avec des doses inférieures de C ont été étudiés, les ETS8 ont été comparés entre les groupes de manière globale par test de Kruskal-Wallis, puis les bras de combinaison C+M ont été comparés au bras de référence C par test de Dunnett.

Pour un nombre identique d'individus, le protocole à un bras présente l'avantage d'obtenir une meilleure puissance statistique qu'un autre protocole, à condition de se comparer à la bonne valeur de référence. Par exemple, avec les ETS8 observés et le scénario d'additivité, la puissance est de 89 % avec ce protocole contre 67 % pour un design à deux bras. Néanmoins, la valeur historique à laquelle on compare la distribution d'ETS8 peut être incorrecte parce que les conditions d'études ou les critères d'inclusion sont différents de l'étude qui a permis d'obtenir la valeur historique. Cela peut conduire à une inflation importante de l'erreur de type 1. Par exemple, si la valeur de référence d'ETS8 choisie était 20 % alors que la valeur correcte était 29 %, cela conduisait à une erreur de type 1 de 57 % (hors de l'intervalle de prédiction à 95 % [3,3 ; 7,3] %). De plus, si la valeur historique est plus élevée que la valeur réelle, cela pouvait conduire à une perte de puissance, par exemple si la valeur de référence choisie était 35 % au lieu de 29 %, la puissance était seulement de 45 % (vs. 89 % avec la valeur correcte). L'essai randomisé à deux bras présente l'avantage important de ne pas avoir à faire d'hypothèse sur la valeur de référence lors de l'analyse des données. Ce design permet de contrôler l'erreur

de type 1 qui était de 6 % dans notre exemple. De plus, dans le scénario de synergie, ce protocole permettait d'obtenir une puissance de 97 % (vs. 100 % pour un design à un bras). Enfin, le protocole à 4 bras peut être intéressant lorsque l'on recherche à sélectionner la dose la plus adéquate, par exemple pour des raisons de toxicité. En revanche, étant donné le plus faible nombre d'individus par bras avec ce protocole et la procédure statistique plus conservatrice, dans le scénario d'additivité, les puissances pour détecter un effet de la combinaison avec la dose forte de C (standard), moyenne et faible étaient respectivement de 20, 13 et 3 %. Dans le cas d'une synergie, les puissances respectives atteignaient 60, 52 et 26 %.

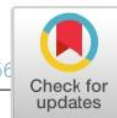
Par ailleurs, certains éléments de ce travail ont montré l'intérêt des modèles mixtes pour l'analyse des données longitudinales de tailles tumorales. En effet, ce type de données est sujet à une erreur de mesure pouvant être importante. Les modèles permettent de prendre en compte ces erreurs de mesure, et ceci permet de réduire la perte de puissance qui leur sont dues. Nous avons observé cette tendance, par exemple, pour le design à quatre bras et le scénario de synergie, la puissance était de 65 % à partir des prédictions basées sur le modèle contre 60 % à partir des données observées.

En conclusion, ce travail a notamment permis de mettre en évidence les avantages et inconvénients de différents designs pour évaluer l'ETS pour un contexte de combinaison de molécule dont l'une est le traitement de référence. Il a fait l'objet d'un article publié dans *Clinical Pharmacology and Therapeutics: Pharmacometrics & Systems Pharmacology*.

2.3.2. Article 3 (publié)

Comparison of various phase I combination therapy designs in oncology for evaluation of early tumor shrinkage using simulations, publié dans *Clinical Pharmacology and Therapeutics: Pharmacometrics & Systems Pharmacology*, 2020

Jérémy Seurat, Pascal Girard, Kosalaram Goteti, France Mentré



ARTICLE

Comparison of Various Phase I Combination Therapy Designs in Oncology for Evaluation of Early Tumor Shrinkage Using Simulations

Jérémy Seurat^{1,*}, Pascal Girard², Kosalaram Goteti³ and France Mentré¹

There is still a lack of efficient designs for identifying the dose response in oncology combination therapies in early clinical trials. The concentration response relationship can be identified using the early tumor shrinkage time course, which has been shown to be a good early response marker of clinical efficacy. The performance of various designs using an exposure–tumor growth inhibition model was explored using simulations. Different combination effects of new drug M and cetuximab (reference therapy) were explored first assuming no effect of M on cetuximab (to investigate the type I error (α)), and subsequently assuming additivity or synergy between cetuximab and M. One-arm, two-arm, and four-arm designs were evaluated. In the one-arm design, 60 patients received cetuximab + M. In the two-arm design, 30 patients received cetuximab and 30 received cetuximab + M. In the four-arm design, in addition to cetuximab and cetuximab + M as standard doses, combination arms with lower doses of cetuximab were evaluated (15 patients/arm). Model-based predictions or “simulated observations” of early tumor shrinkage at week 8 (ETS8) were compared between the different arms. With the same number of individuals, the one-arm design showed better statistical power than other designs but led to strong inflation of α in case of misestimated reference for ETS8 value. The two-arm design protected against this misestimation and, with the same total number of subjects, would provide higher statistical power than a four-arm design. However, a four-arm design would be helpful for exploring more doses of cetuximab in combination with M to better understand the interaction.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Currently, there is no standard for design choice in early clinical studies of combinations treatments in immuno-oncology.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ These clinical trial simulations compared designs for the first time in the context of a fixed dose of one agent assumed to be optimal while testing different doses of another based on early tumor shrinkage.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ Statistical tests performed on individual model-based early tumor shrinkage at week 8 values and the one-arm design indicate a better power than two-arm or four-arm designs, but imply strong assumptions about the historical reference value, leading to strong inflation of type I error in the case of underestimating the reference. Choosing a

two-arm or a four-arm design depends on the objective of the study: a two-arm design is preferable to a four-arm design to achieve a good statistical power, but a four-arm design allows better exploration of the combination and better dose selection.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

☑ These simulations show that, in early-phase oncology combination studies (with early tumor shrinkage as the outcome), the choice of a one-arm design (i.e., without a comparator arm) is not appropriate unless the conditions of the clinical trial are similar to those of the reference treatment or because of ethical reasons. This implies strong assumptions about the historical value, and the trial could lead to a wrong conclusion if the historical value is no longer relevant.

Early-phase studies are critical in oncology drug development. These trials are frequently designed to evaluate the overall response as well as toxicity and pharmacokinetics. Go/no go decisions and dose recommendations for further studies are based on the results of phase I studies.

Therefore, the design choice is crucial in the early phase.¹ There is growing interest in the use of combination therapies in oncology, especially immuno-oncology.² The combination of several agents could have a synergistic effect and enhance the antitumor activity. However, toxicities are

¹Université de Paris, INSERM, IAME, F-75006 Paris, France; ²Merck Institute for Pharmacometrics, Merck Serono S.A, Lausanne, Switzerland; ³EMD Serono R&D Institute, Billerica, Massachusetts, USA. *Correspondence: Jérémy Seurat (jeremy.seurat@inserm.fr)
Received: June 26, 2020; accepted: September 21, 2020. doi:10.1002/psp4.12564

frequently observed in these studies, especially when chemotherapies are involved.³ In the past two decades, several new targeted therapies, such as monoclonal antibodies, with efficacy in cancer subpopulations and limited adverse effects compared with traditional therapies have reached the market.⁴ These drugs have recently been tested in many combination trials in addition of chemotherapy/radiotherapy.^{5,6} However, the designs of clinical trials investigating combinations of immune-oncology agents are often based on empiricism without clear assumptions about the drug interactions. Therefore, the design choice in combination trials could be challenging, especially when clinical information is only available for single agent given as monotherapy. Recommendations have been made concerning the design of early studies with a combination of targeted anticancer therapies.⁷ The importance of the assumed pharmacodynamic drug interaction and the usefulness of biomarkers are highlighted in these recommendations. More recent recommendations insist on the importance of including drug concentrations in the analysis and understanding the pharmacokinetic/pharmacodynamic relationships.⁸ The importance of including biomarker-driven objectives in phase I trials has also been shown.⁹

Among clinical trial end points widely used in oncology, overall survival is universally accepted but involves large studies with long follow-up, which is not compatible with early-phase studies. End points based on tumor assessment can be used. The Response Evaluation Criteria in Solid Tumors (RECIST) criteria, widely adopted in oncology clinical trials including on targeted therapies, consist of four categories that depend on the evolution of tumor size (TS).^{10,11} However, in the field of modeling and simulation it seems preferable to use the continuous TS data when available rather than RECIST as this would result in a loss of information.^{12,13} Furthermore, the link between the early tumor shrinkage (ETS) and overall survival or progression free-survival has been proposed as an early efficacy marker based on several studies.¹⁴⁻¹⁶

Nonlinear mixed effect models (NLMEM) are increasingly used to support drug decision making, especially in early-phase trials.¹⁷ In oncology, these models can be used to capture the pharmacokinetics of antitumoral agents or biomarker kinetics.¹⁸ NLMEM can describe tumor dynamics, and several tumor growth models are found in the literature such as Gompertz¹⁹ or Simeoni.²⁰ Tumor growth inhibition (TGI) models including a drug induced decrease in TS are used for various types of cancer and antitumoral agents, including targeted therapies.^{15,21} TS (or other biomarkers) can also be linked to survival by a joint model.²² However, before conducting a trial and modeling the data, it is crucial to choose an appropriate design to obtain good precision of the parameter estimates and thus precise results. For this purpose, model-based clinical trial simulation (CTS) can be performed to evaluate and compare designs.

In simulations, we used a combination of candidate drug M given on top of cetuximab, a monoclonal antibody targeting the epidermal growth factor receptor of tumor cells²³ that has been approved for metastatic colorectal cancer with wild-type RAS and usually given in combination with

chemotherapy. Several ongoing or completed studies indicate that new combinations including cetuximab could be promising for different cancers.^{24,25} The aim of this study was to compare by CTS several combination designs in the treatment of refractory colorectal cancer with different assumptions about the interaction between cetuximab and M. The performances, type I error (α), and power of several designs were compared to test the superiority of the combination treatment to cetuximab alone based on ETS at week 8 (ETS8). Another objective was to compare, by modeling and simulation of exposure-TGI, the test performances on predicted ETS8 obtained from the model to the observed ETS8.

METHODS

The different models used to perform the simulations were the model of pharmacokinetic exposure during the first 8 weeks of therapy for each drug and the TGI model. We studied three different designs (one arm, two arms, and four arms) in three different scenarios: no effect of M, additivity, and synergy. Individual ETS8 values were computed as predictions after global model fitting, observations, or true values.

Models

Cetuximab and drug M pharmacokinetic exposures. Cetuximab has been described by various pharmacokinetic models.^{26,27} A significant influence of body surface area (BSA) on clearance of cetuximab has been shown.²⁸ Therefore, it is recommended to adjust the dose of cetuximab ($Dose_C$) on BSA.²⁹ We assumed a simple model with independent clearance with dose, as found in a previous study in patients with metastatic colorectal cancer.³⁰ Therefore, cetuximab exposure expressed as area under the curve (AUC_C) integrated over 2 weeks (between two administrations), in $mg \times week/L$, was simulated as in Eq. 1.

$$AUC_{C:2\text{ weeks}} = \frac{Dose_C(mg/m^2) \times BSA(m^2)}{CL_C(L/week) \times (BSA/1.85)^{0.75}}, \quad (1)$$

where CL_C is the cetuximab clearance value (CL_C) presented in **Table 1**, including BSA effect standardized by 1.85 (i.e., the median BSA that was expected before the clinical trial), and BSA was simulated according to a lognormal distribution with a median of 1.75 and m^2 and an interindividual variability of 10% according to derived BSA distribution in previous cetuximab studies.³¹⁻³³

The current maximal dosing regimen for cetuximab given in combination with another treatment is 500 mg per m^2 every 2 weeks. This dosing regimen has a comparable activity and safety to the standard dosing regimen.³⁴ A dosing regimen of 400 and 200 mg per m^2 every 2 weeks were also considered in case of combination with drug M.

Concerning the drug M, its area under the curve (AUC_M) was integrated over 2 weeks (between two administrations) and simulated as in Eq. 2:

$$AUC_{M:2\text{ weeks}} = \frac{Dose_M}{CL_M}, \quad (2)$$

Table 1 Parameter values for area under the curve models of cetuximab and drug M and for the tumor growth inhibition model

Parameter (unit)	μ	ω^2		
Pharmacokinetics models				
CL _C (L.week ⁻¹)	3.9	0.0025		
CL _M (L.week ⁻¹)	5.0	0.01		
Tumor growth inhibition model				
TS0 (mm)	100 ^a	0.5 ^a		
KS (week ⁻¹)	0.001	—		
KD _{SOC} (week ⁻¹)	0.015	1.5		
KD _C (L × mg ⁻¹ .week ⁻²)	0.00025 ^a	1.5 ^a		
KD _M (L × mg ⁻¹ .week ⁻²)	0.00025 ^a	1 ^a		
KR (week ⁻¹)	0.2	1		
Interaction parameters				
	Additivity	Synergy	Additivity	Synergy
INT	0	2 ^a	—	0.16 ^a
INT ₅₀ (mg × week/L)	—	75	—	—
Correlations				
Cov(η_{KR}, η_{KDC})			1 ^a ($\rho = 0.82$)	
Residual error				
σ_{prop}^2			0.023 ^a	

μ are the fixed effects, ω^2 the variance of random effects η , and σ^2 the variance of residuals.

CL_C, cetuximab clearance value; CL_M, clearance of drug M; Cov, covariance; INT, interaction parameter; INT₅₀, exposure needed for 50% of interaction maximal effect; KD_C, effect of drug cetuximab; KD_M, effect of drug M; KD_{SOC}, standard of care effect; KR, constant rate of resistance development; KS, natural growth of TS; L.week⁻¹, Litre.week⁻¹; η , random effects; TS0, tumor size at baseline.

^aEstimated parameter.

where CL_M is the clearance of drug M (CL_M) presented in **Table 1**.

Tumor size. Tumor size is defined as the sum of the longest diameters of target lesions that are simulated through a TGI model.¹⁵ The concentration of anticancer antibodies such as cetuximab is associated with the response rate.²⁷ A previous TGI model was developed based on clinical data from patients affected by metastatic colorectal cancer.²⁸ In this analysis, the TS dynamics was described by the ordinary differential equation (Eq. 3) linking TS and both drug exposures (AUC_C and AUC_M).

$$\begin{aligned} \frac{dT_S}{dt} &= KS \times TS - KD \times e^{-KR \times t} \times TS \\ KD &= KD_{SOC} + KD_C \times AUC_C + KD_M \times AUC_M \times \left(1 + \frac{INT \times AUC_C}{INT_{50} + AUC_C}\right) \quad (3) \\ TS(0) &= TS0 \end{aligned}$$

In this model (see Supplementary file), KS is the natural growth of TS, KD includes the different pooled drug effects on tumor shrinkage, KR represents the constant rate of resistance development, and TS0 is the TS at baseline. KD is the sum of KD_{SOC} (for standard of care, chemotherapy), KD_C × AUC_C (for cetuximab, with KD_C the effect of cetuximab), and KD_M × AUC_M × $\left(1 + \frac{INT \times AUC_C}{INT_{50} + AUC_C}\right)$ (for drug M, with KD_M the effect of M).

According to Eq. 3, the effect of drug M is the sum of its proper effect and its interaction with cetuximab. This interaction model is inspired by the global pharmacodynamics interaction model.³⁵ Depending on the sign of the interaction parameter (INT), this model is able to capture different types of interactions. A positive INT represents synergism, a negative INT represents antagonism, and if this parameter is null, the effects of both drugs are additive. Parameter

values of this model are given in **Table 1** based on previous reported analysis on cetuximab (excepted KD_M and interaction parameters) in the treatment of colorectal cancer.²⁸ The standard of care effect (KD_{SOC}) includes the placebo response. KD_M is selected from a sensitivity analysis to obtain plausible tumor shrinkage. Two interaction scenarios are investigated: additivity in which the two compounds act independently and synergy in which the cetuximab exposure increases the drug M effect. The difference between these two scenarios in terms of median TS evolution from 0 to 8 weeks is illustrated in **Figure S1**.

Let i denote the i^{th} individual ($i = 1, \dots, N$) and j the j^{th} TS measurement of an individual. The statistical model for the TS observation TS_{ij} in individual i at time t_{ij} is given by Eq. 4.

$$TS_{ij} = f(t_{ij}, \theta_i) + \varepsilon_{ij}, \quad (4)$$

where f is the TGI model described in Eq. 3, θ_i is the vector of individual parameters, and ε_{ij} is the residual errors of mean 0 and of variance proportional to the TS: $\text{Var}(\varepsilon_{ij}) = f(t_{ij}, \theta_i)^2 \times \sigma_{prop}^2$. The random effect model is exponential as $\theta_i = \mu \times \exp(\eta_i)$, where μ is the fixed effect and η_i the random effect, except for INT in which the random effect model is additive: $\theta_i = \mu + \eta_i$. The AUCs of both cetuximab and M are simulated assuming a constant clearance and used as regressors in the TGI model.

Simulation settings

Designs. We studied three types of design—one arm, two arms, or four arms—considering the same total number of individuals ($N = 60$). Simulated TS are performed at predose (baseline) and weeks 2, 4, 6, and 8. The inclusion criterion was a baseline TS > 20 mm. Dropout, incomplete adherence, missing TS measurements, and protocol

deviations were not considered. The therapy for all patients included a standard of care, which is omitted next for simplification. Simulations were performed using R software (R Foundation for Statistical Computing, Vienna, Austria) version 3.4.3., specifically, packages *mvtnorm* to generate individual parameters³⁶ and *deSolve* to simulate longitudinal data.³⁷ Simulated TS < 10 mm were censored and considered to be below the limit of quantification. The dosing regimens of cetuximab and drug M for each possible design are described in the following paragraphs. The dose of M was assumed to be fixed. For each design and each interaction scenario, 500 data sets were simulated.

One-arm: M + C500. In the one-arm design, the $N = 60$ patients were allocated to a combination M + C500: drug M and cetuximab. The dosing regimen was fixed for each drug: cetuximab at 500 mg/m² every 2 weeks (C500) and drug M at its standard dose.

Two-arm: C500/M + C500. In the two-arm design, the $N = 60$ patients were randomized among the two possible arms: cetuximab “alone” (C500) or M + C500, resulting in 30 patients in each arm. Typical simulated data for patients in the two arms for the additivity and synergy scenarios are presented in **Figure S2**.

Four-arm: C500/M ± C200/M ± C400/M + C500. In the four-arm design, two combination arms with lower doses of cetuximab were considered (M + C400 and M + C200) in addition to C500 and M + C500. In the lower dose arms, the dose of C was 400 mg/m² (C400) or 200 mg/m² every 2 weeks. Each arm included 15 patients.

Scenarios of the effect of M + C. Three scenarios concerning the effect of M and the interaction between M and C were considered. To investigate the type I error of tests (superiority of M + cetuximab vs. cetuximab alone), a first scenario without an effect of drug M (i.e., with $KD_M = 0$) was studied. In the other two scenarios, an effect of M ($\mu_{KDM} = 0.00025$) was implemented. These two scenarios differ in the interaction between M and cetuximab: additivity ($\mu_{INT} = 0$) or synergy ($\mu_{INT} = 2$).

Analysis

Predictions, observations, and true individual ETS8 values. For each individual i , ETS8 is expressed as a percentage and calculated as in Eq. 5.

$$ETS8_i = \frac{(TS_i(0) - TS_i(8))}{TS_i(0)} \times 100, \quad (5)$$

where $TS_i(0)$ and $TS_i(8)$ are individual TS at baseline and after 8 weeks of treatment, respectively. Predicted (IPRED), observed (OBS), and true (TRUE) ETS8 are calculated using predictions, observations, and true individual TS at both week 0 and week 8, respectively, as described in the next paragraphs.

Individual predictions (IPRED) were obtained from model-based parameters. Parameters with a superscript letter in **Table 1** are estimated based on the TGI model (Eq. 1) by fitting

the simulated data using the SAEM (Stochastic Approximation Expectation-Maximization) algorithm³⁸ in MONOLIX 2018R2 software (Lixoft, Antony, France). MONOLIX runs were performed with five Markov chains, a minimum of 500 iterations during the exploratory phase and a minimum of 200 iterations during the smoothing phase. An adequate method is used to account for below the limit of quantification data.³⁹ TS individual predictions at baseline and week 8 were obtained using the mode of empirical Bayes estimates of the individual parameters. In the synergy model, the μ_{INT} and its interindividual variability ω_{INT} were estimated, whereas in the additivity model they were fixed to 0. Simulated data without an effect of M were refitted for both the additivity and synergy models to evaluate the type I error for each assumed interaction.

Individual simulated TS observations (OBS) were directly calculated from the simulated TS values plus residual error while true individual TSs (TRUE) were simply derived from individual parameters without residual error. It is important to realize that true values are not observed in a real clinical trial. An example of OBS, IPRED, and TRUE TSs in one individual is illustrated in **Figure S3**.

Design evaluation. The performances of each design were evaluated and compared in terms of type I error and power to detect the superiority of the combination (drug M + cetuximab) to cetuximab based on the ETS8 values (OBS, IPRED, and TRUE). The statistical tests described in the next paragraphs were two-sided using $\alpha = 5\%$.

For the one-arm design, as all the individuals were allocated to the combination arm, individual ETS8 values were compared with a reference ETS8 for cetuximab alone from historical data using a one-sample Wilcoxon test. Different scenarios were investigated concerning the reference in which the historical value is either not valid (lower or higher than true, due, e.g., to standard of care changes or differences in patient cohorts) or true for this reference ETS8. In this study, the reference value was 29%, (i.e., the median of all individual simulated ETS8 values with C500).

For the randomized trial with two arms, C500 and cetuximab + M, a two-sample Wilcoxon test was performed to compare the ETS8 values of individuals from each arm.

In the four-arm design, a global Kruskal-Wallis test was performed first to compare the ETS8 values between the four arms. If a difference between at least two arms was detected by this global test, a Dunnett test was performed to compare each combination arm to the reference C500 alone using a Benjamini-Hochberg procedure to account for multiplicity of tests.⁴⁰

From the data sets simulated without an effect of drug M ($KD_M = 0$, i.e., under H_0), the type I error was evaluated as the proportion of trials for which H_0 is rejected and compared to the nominal value of 5%. The 95% prediction interval for proportion π of H_0 rejection (type I error or power) is given by binomial law⁴¹ ((3.3%, 7.3%) with 500 data sets). From the data sets simulated with an effect of drug M (additivity or synergy scenario, i.e., under H_1), the power of the tests was evaluated as the proportion of trials for which H_0 is rejected.

RESULTS

First, we checked whether model-based ETS8 values (IPRED) were overpredicting or underpredicting the true values. Overall, ETS8 are distributed accurately around the identity line for each design and under both additivity and synergy scenarios. However, a trend of overpredicted values for ETS8 < 25% is shown, although it seems to not be influenced by the scenario, design, or treatment arm (Figure S4).

One-arm: M + C500

For the one-arm design, the ETS8 values for patients were compared to an ETS8 value for current reference therapy obtained from historical data. Distributions of individual ETS8 (OBS, IPRED, and TRUE) are presented in Figure 1 in the cases of additivity and synergy. The median of OBS, IPRED, and TRUE ETS8 values were close to each other in each scenario. The observed median ETS8 was 43% for additivity and 55% for synergy, whereas the reference ETS8 for cetuximab alone was 29%. According to our model and simulation settings, assuming additivity, almost three of four patients had greater tumor shrinkage at week 8 than those of reference treatment. These boxplots enable us to see the interindividual variability in ETS8. As predictions account for residual error, the 5th to 95th percentile of ETS8 is narrower for IPRED (from 23% to 78% for additivity and from 28% to 92% for synergy) than OBS (from 3% to 78% for additivity and from 13% to 89% for synergy). In both scenarios, the distribution of predictions seems to be more representative of true values than observations.

With this one-arm design, as the ETS8 distribution is only known for M + C500, a one-sample Wilcoxon test was performed (Table 2). Using observed data and assuming a reliable ETS8 reference value under C500 (in our case 29%), the type I error is controlled (5.6%) as it lies in the prediction interval around the nominal value $\alpha = 5\%$. However, the type I error is not controlled using model-based predictions (12.6% for additivity and 13.2% for synergy), which is partly attributed to a lack of identifiability between the M and cetuximab effect. Overall, assuming a reliable reference, good power (at least 89%) is reached using either OBS or IPRED.

In this kind of design, for multiple possible reasons, the assumed reference value from historical data could be

wrong. Therefore, we investigated the type I error and power with lower (25% and 20%) and higher (35% and 40%) ETS8 reference values. Performing a two-sided Wilcoxon test, the type I error was inflated in cases using either OBS or TRUE. For example, if we compare the ETS8 distribution to 20% instead of 29%, we wrongly conclude to superiority of the combination in more than the half of the CTS using observations or predictions. Moreover, assuming a higher reference value than true leads to a lack of power. For example, in the additivity scenario, assuming a reference value of 35% leads to a power of 45.2% using observations (instead of 89.0% with the real reference value).

Two-arms: C500/M + C500

For the randomized two-arm design, the distribution of individual ETS8 (OBS, IPRED, and TRUE) values in each arm is shown in Figure 2 for additivity and synergy. As with the one-arm design, the median OBS, IPRED, and TRUE ETS8 values are close to each other in each arm and each scenario. For C500, the median ETS8 is ~ 29% (i.e., the reference value in the one-arm design) and for M + C500, the median ETS8 is ~ 43% with additivity and ~ 55% with synergy. Notably, with C500, more than 1 of 10 patients presented a negative observed ETS8 (i.e., increasing TS between weeks 0 and 8), whereas none of the individuals had a true negative ETS8 under this therapy with our simulation settings, which is captured well using predicted ETS8. The IPRED distribution of C500 seems to be similar between the two scenarios. This indicates that the estimated cetuximab specific effect is not influenced by the model or the scenario, as in the simulated data, and allows for comparison between the scenarios. As with the individuals given M + C500 in the one-arm design, the variability in ETS8 is considerable with C500 in the two-arm designs, especially in OBS (the 5th to 95th percentile extends from -12% to 65%). Therefore, the difference in ETS8 distribution between the two arms is difficult to distinguish in the case of additivity. The IPRED distribution is narrower and closer to TRUE distribution than OBS.

With this design, comparing ETS8 distributions between the two arms by performing two-sample Wilcoxon tests, no assumption on ETS8 reference value is required (Table 3). Therefore, choosing a two-arm design instead of a one-arm design avoids a potential inflation of the type I error. In both

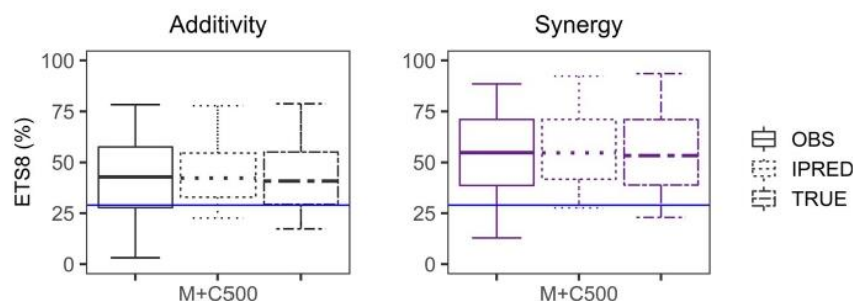


Figure 1 Boxplots of early tumor shrinkage at week 8 (ETS8) from observations (OBS; solid boxes, left subgroup for each treatment), predictions (IPRED; dotted boxes, middle subgroup for each treatment), and true values (TRUE; dashed boxes, right subgroup for each treatment) in the one-arm design (combination M + C500). The blue horizontal line is the true median ETS8 (29%) of the historical reference treatment. Whiskers indicate 5th to 95th percentiles.

Table 2 Type I error and power of tests in one-arm design

Theoretical median of ETS8 from historical data (%)	Additivity						Synergy						
	Type I error (%)			Power (%)			Type I error (%)			Power (%)			
	OBS	IPRED	TRUE	OBS	IPRED	TRUE	OBS	IPRED	TRUE	OBS	IPRED	TRUE	
Lower than true	20	56.6	79.4	68.6	100.0	100.0	100.0	56.6	60.2	68.6	100.0	100.0	100.0
	25	13.2	34.2	18.0	98.6	99.6	99.8	13.2	34.2	18.0	100.0	100.0	100.0
True reference	29	5.6	12.6	6.0	89.0	97.6	93.6	5.6	13.2	6.0	100.0	100.0	100.0
Higher than true	35	27.4	31.6	37.4	45.2	70.6	51.0	27.4	32.2	37.4	99.4	100.0	99.8
	40	74.8	75.0	82.6	9.8	25.0	10.6	74.8	74.8	82.6	90.2	98.0	93.0

The type I error and power from clinical trial simulations are given as the observed proportion of significant tests performed on early tumor shrinkage at week 8 (ETS8) from 500 simulated clinical trials. Under H_0 , the 95% prediction interval from binomial distribution is 3.3%–7.3% for 500 simulations. For the one-arm design, the number of individuals by data set is $N = 60$, and all individuals are allocated to the combination arm. The ETS8 distribution (of observations (OBS), predictions (IPRED), or true values (TRUE)) is compared with a theoretical median of ETS8 under cetuximab alone, performing a one-sample Wilcoxon test (two-sided, $\alpha = 5\%$). Different possible theoretical median of ETS8 values (true or not) are investigated. Two possible drug interactions are considered: additivity and synergy (see Models section).

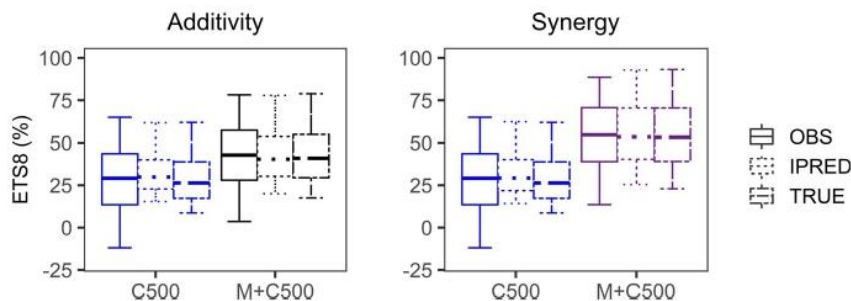


Figure 2 Boxplots of early tumor shrinkage at week 8 (ETS8) from observations (OBS; solid boxes, left subgroup for each treatment), predictions (IPRED; dotted boxes, middle subgroup for each treatment), and true values (TRUE; dashed boxes, right subgroup for each treatment) in the two-arm design. Whiskers indicate 5th to 95th percentiles. C500 is cetuximab alone vs. the combination treatment (M + C500).

additivity and synergy scenarios and using either observed or predicted ETS8, the type I error is well controlled as it lies in the 95% prediction interval (3.3%–7.3%). Compared with the previous design, there is a decrease in power in the additivity scenario (using OBS: 66.6% vs. 89.0% in one-arm design). This result is explained by having only 30 patients per arm (vs. 60 in the one-arm design). Nevertheless, this design remains powerful in the case of synergy (using OBS: 97.4% vs. 100% in the one-arm design). Moreover, in each scenario, using the model-based ETS8 value leads to a slight increase in power compared with using the raw data. For example, the power in the additivity scenario was 69.2% using IPRED.

Four-arms: C500/M + C500/M + C400/M + C200

For the four-arm design, the distributions of individual ETS8 values (OBS, IPRED, and TRUE) in each of the arms are presented in **Figure 3** for additivity and synergy. The IPRED distribution seems to be closer to TRUE than OBS, especially for extreme percentiles, except in the C500 arm. As expected, the distributions of C500 and M + C200 are close to each other for additivity. Moreover, in synergy, the ETS8 distribution is close between the two combination arms with the higher doses of cetuximab.

Type I error and power are presented in **Table 4**. Performing a global Kruskal–Wallis to detect a difference between the ETS8 distributions of the four arms, more power was

Table 3 Type I error and power of tests in two-arm design

	Additivity		Synergy	
	Type I error (%)	Power (%)	Type I error (%)	Power (%)
OBS	6.2	66.6	6.2	97.4
IPRED	6.2	69.2	7.0	98.6
TRUE	6.2	89.0	6.2	99.8

The type I error and power from clinical trial simulations are given as the observed proportion of significant tests performed on early tumor shrinkage at week 8 (ETS8) from 500 simulated clinical trials. Under H_0 , the 95% prediction interval from binomial distribution is 3.3%–7.3% for 500 simulations. For the two-arms design, 30 patients are allocated to the combination arm, and 30 are allocated to cetuximab alone. The ETS8 distributions (of observations (OBS), predictions (IPRED), or true values (TRUE)) are compared between the two arms, performing a two-sample Wilcoxon test (two-sided, $\alpha = 5\%$). Two possible drug interactions are considered: additivity and synergy (see Models section).

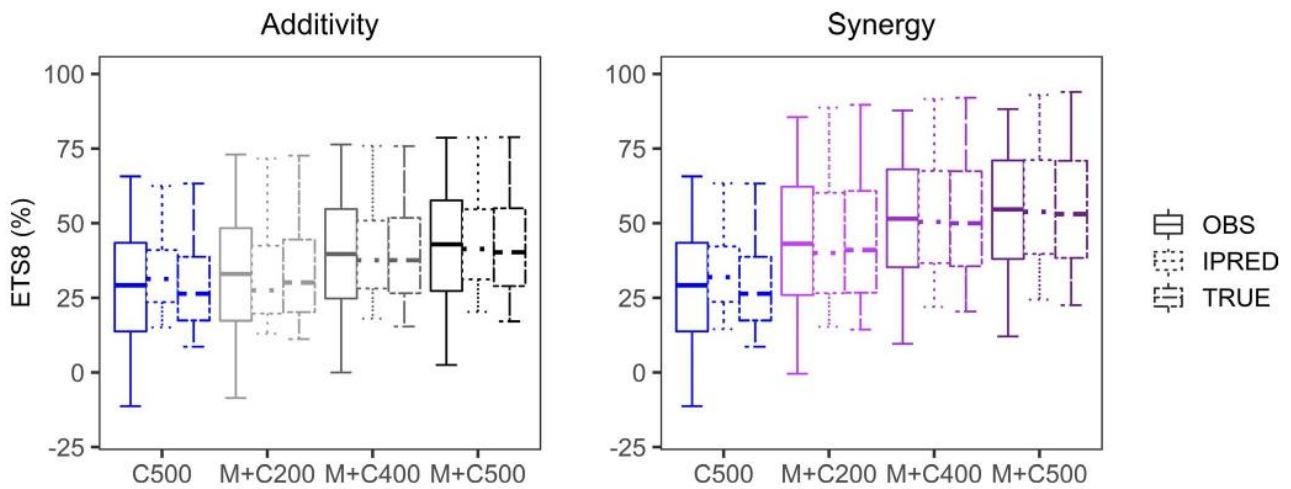


Figure 3 Boxplots of early tumor shrinkage at week 8 (ETS8) from observations (OBS; solid boxes, left subgroup for each treatment), predictions (IPRED; dotted boxes, middle subgroup for each treatment), and true values (TRUE; dashed boxes, right subgroup for each treatment) in the four-arm design. Whiskers indicate 5th to 95th percentiles. C500 is cetuximab alone, and M + C200, M + C400, and M + C500 are combination treatments.

obtained in the synergy scenario than the additivity scenario. For example, using observed ETS8, the power was 69.4% in synergy vs. 30.2% in additivity. If the global test was significant, a Dunnett test was performed, comparing the ETS8 distribution in each of the three combination arms to those of the reference arm C500. Under the null hypothesis, if we compare the distribution of ETS8 for C500 and M + C500, the Dunnett test was conservative compared with a single Wilcoxon test, the type I error using observations was 2% with this procedure in the four-arm design vs. 6.2% for the Wilcoxon test in the two-arm design. Therefore, in the additivity scenario, the power to detect the superiority of M + C500 vs. C500 decreased to 19.8% using observations and 25.6%

using predictions (vs. 66.6% and 69.2% in two-arm design, respectively), which is partly explained by the lesser number of patients per arm (15 vs. 30). Nevertheless, in synergy and using model-based ETS8, we achieved 65.4% power to detect the superiority for M + C500 and 52% for M + C400. Thus, under certain conditions, this design is interesting for exploring other doses in the combination of two drugs.

DISCUSSION

This work compares designs for early clinical combination trials in oncology in the context of a fixed dose of one agent and different doses of another with the same number of

Table 4 Type I error and power of tests in four-arm design

	Additivity			Synergy		
Power of Kruskal–Wallis test (%)						
OBS	30.2			69.4		
IPRED	52.0			76.6		
TRUE	44.0			87.8		
Type I error of Dunnett test (%) C500 vs. M + C500						
OBS	1.6			1.6		
IPRED	2.0			2.0		
TRUE	2.4			2.4		
	C500 vs. M + C200	C500 vs. M + C400	C500 vs. M + C500	C500 vs. M + C200	C500 vs. M + C400	C500 vs. M + C500
Power of Dunnett test (%)						
OBS	3.0	13.0	19.8	26.4	51.8	60.2
IPRED	5.4	13.2	25.6	18.6	52.0	65.4
TRUE	4.0	19.8	29.2	34.0	70.4	79.0

The type I error and power from clinical trial simulations are given as the observed proportion of significant tests performed on early tumor shrinkage at week 8 (ETS8) from 500 simulated clinical trials. For the four-arms design, 15 patients are allocated to each arm. The ETS8 distributions (of observations (OBS), predictions (IPRED), or true values (TRUE)) are compared between the arms, performing first a global Kruskal–Wallis test (two-sided, $\alpha = 5\%$), then, if the global test is significant, a Dunnett test where distribution of each combination arm (M + C200, M + C400, and M + C500) is compared with those of cetuximab alone (C500). Two possible drug interactions are considered: additivity and synergy (see Models section).

subjects and based on ETS. The one-arm design leads to inflation of type I error in the case of a misestimated reference value, as reported previously.⁴² Therefore, before the analysis of data from a one-arm design, we should ensure that the conditions of the study are similar to the reference study. With the same number of subjects and an accurate reference value, the power of a one-arm design is higher than that of a two-arm design. However, to avoid the risk of an inaccurate reference value, a two-arm randomized design including an arm with the reference therapy is preferable. Such a design conducts to control the type I error. Another possibility is to choose a four-arm design with different possible dosing regimens. Although with the same total number of patients a two-arm design leads to a greater power of tests, a four-arm design offers more accurate dose selection.

Our work aimed to show simulations in an ideal setting in order to facilitate the interpretation of the results. In these simulations, we did not account for possible incomplete adherence, missing data, protocol deviations (e.g., timing of TS measurements), or dropout. It would be interesting to incorporate these design departures in future simulations and analyses to mimic a real clinical study.⁴³ Nevertheless, most of these departures are not preponderant in the first 8 weeks of a trial and therefore should not change drastically our simulation results. In this study, exposure to cetuximab was simulated from a previous pharmacokinetics model²⁸ that included a BSA effect on clearance but other covariates, such as normal fat mass, would be more physiological.⁴⁴ Moreover, the allometric coefficient of 0.75 (used for the BSA effect in the model) would be more appropriate with the normal fat mass. In our study, the clearance of M was not influenced by any covariate, but the effect of body weight or the normal fat mass could have been considered. In the model that we used, we assumed linear effects in the exposure ranges of cetuximab and M based on previous models in which the maximal effect (E_{max}) could not be estimated. However, it would be relevant to explore more mechanistic models and assume nonlinear effects, as is the case for the interaction between cetuximab and M in the model. The TGI model involves a global resistance parameter KR. It would be relevant to associate different resistance parameters with each drug and its exposure.

Using NLMEM to analyze longitudinal data is a promising tool, especially when the outcome is based on TS. Tumor measurements are notably influenced by scan variability and the individual reviewing the image, which may result in noise in the data⁴⁵ and a loss of power. The NLMEM accounts for the information from the longitudinal data, which were five TS measurements in our study, and allows estimation of the residual error, including the noise in the data. Therefore, performing tests based on model-based predictions can lead to better power than observations. For example, in our two-arm design with the additivity scenario, the power using model-based ETS8 was 69.2% vs. 66.6% using observations.

In this study, ETS was chosen as outcome. We could also dichotomize ETS as responder or nonresponder outcome. Different thresholds can be chosen to separate the patients depending on their ETS.⁴⁶ We can dichotomize the

subjects as responders if $ETS8 \geq 30\%$ or nonresponder if $ETS8 < 30\%$, which corresponds roughly to a complete or partial response vs. stable or progressive disease according to the RECIST criteria. With this binary outcome, in our two-arm design and additivity scenario, the power of Fisher's exact test was 46.4% instead of 69.2% with the Wilcoxon test of continuous model-based ETS8 values.

Guidelines highlight the importance of design choice for an early-phase combination trial in oncology and that assumptions and knowledge about pharmacokinetic/pharmacodynamic interactions should be taken into account to select the most effective trial design.⁴⁷ There is no standard design for combination studies, and designs that include randomization helps to define the effect of the combination vs. one or both of the individual components.¹

We plan to extend this work to perform model-based adaptive two-stage designs^{48,49} using the Fisher information matrix to optimize the second stage of the study⁵⁰ in which different dose arms could also be added or dropped at the end of the first stage.

Supporting Information. Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (www.psp-journal.com).

Acknowledgments. The authors would like to thank Thu Thuy Nguyen and Vishnu Dutt Sharma who contributed to this project. They also thank Hervé Le Nagard and Lionel de la Tribouille for the use of the CATBioMed calculus facility.

Funding. This work was supported by Merck KGaA, Darmstadt, Germany.

Conflict of Interest. Pascal Girard and Kosalaram Goteti are employed by Merck Institute for Pharmacometrics and EMD Serono R&D Institute, Inc., respectively, an affiliate of Merck KGaA, Darmstadt, Germany. All other authors declared no competing interests for this work.

Author Contributions. J.S., P.G., K.G., and F.M. wrote the manuscript. J.S., P.G., K.G., and F.M. designed the research. J.S., P.G., K.G., and F.M. performed the research. J.S., P.G., K.G., and F.M. analyzed the data.

Disclaimer. As Editor-in-Chief of *CPT: Pharmacometrics and Systems Pharmacology*, France Mentré was not involved in the review or decision process for this article.

1. LoRusso, P.M., Boerner, S.A. & Seymour, L. An overview of the optimal planning, design, and conduct of phase I studies of new therapeutics. *Clin. Cancer Res.* **16**, 1710–1718 (2010).
2. Ledford, H. Cocktails for cancer with a measure of immunotherapy. *Nature* **532**, 162–164 (2016).
3. Sandler, A. et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N. Engl. J. Med.* **355**, 2542–2550 (2006).
4. Huang, J. et al. Emerging trends in us oncological approvals: a 13-year review (1999–2011). *Drug Inf. J.* **46**, 344–357 (2012).
5. Cunningham, D. et al. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* **351**, 337–345 (2004).
6. Sullivan, R.J. et al. Atezolizumab plus cobimetinib and vemurafenib in BRAF-mutated melanoma patients. *Nat. Med.* **25**, 929–935 (2019).
7. Seymour, L.K. et al. Design and conduct of early clinical studies of two or more targeted anticancer therapies: recommendations from the task force on methodology

- for the development of innovative cancer therapies. *Eur. J. Cancer Oxf. Engl. 1990* **49**, 1808–1814 (2013).
8. Smoragiewicz, M. et al. Design and conduct of early clinical studies of immunotherapy: recommendations from the task force on Methodology for the Development of Innovative Cancer Therapies 2019 (MDICT). *Clin. Cancer Res.* **26**, 2461–2465 (2020).
 9. Paller, C.J. et al. Factors affecting combination trial success (FACTS): investigator survey results on early-phase combination trials. *Front. Med.* **6** (2019). <https://doi.org/10.3389/fmed.2019.00122>.
 10. Eisenhauer, E.A. et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).
 11. Litière, S. et al. RECIST 1.1 for response evaluation apply not only to chemotherapy-treated patients but also to targeted cancer agents: a pooled database analysis. *J. Clin. Oncol.* **37**, 1102–1110 (2019).
 12. Sharma, M.R., Maitland, M.L. & Ratain, M.J. RECIST: no longer the sharpest tool in the oncology clinical trials toolbox—point. *Cancer Res.* **72**, 5145–5149 (2012), discussion 5150.
 13. Vera-Yunca, D. et al. Machine learning analysis of individual tumor lesions in four metastatic colorectal cancer clinical studies: linking tumor heterogeneity to overall survival. *AAPS J.* **22**, 58 (2020).
 14. Wang, Y. et al. Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clin. Pharmacol. Ther.* **86**, 167–174 (2009).
 15. Claret, L. et al. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J. Clin. Oncol.* **27**, 4103–4108 (2009).
 16. Colloca, G.A., Venturino, A. & Guarneri, D. Early tumor shrinkage after first-line medical treatment of metastatic colorectal cancer: a meta-analysis. *Int. J. Clin. Oncol.* **24**, 231–240 (2019).
 17. Kimko, H. & Pinheiro, J. Model-based clinical drug development in the past, present and future: a commentary. *Br. J. Clin. Pharmacol.* **79**, 108–116 (2015).
 18. Petrylak, D.P. et al. Evaluation of prostate-specific antigen declines for surrogacy in patients treated on SWOG 99–16. *J. Natl. Cancer Inst.* **98**, 516–521 (2006).
 19. Gompertz, B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trans. R. Soc. Lond.* **115**, 513–583 (1825).
 20. Simeoni, M. et al. Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Res.* **64**, 1094–1101 (2004).
 21. Bonate, P.L. & Suttle, A.B. Modeling tumor growth kinetics after treatment with pazopanib or placebo in patients with renal cell carcinoma. *Cancer Chemother. Pharmacol.* **72**, 231–240 (2013).
 22. Tardivon, C. et al. Association between tumor size kinetics and survival in patients with urothelial carcinoma treated with atezolizumab: implication for patient follow-up. *Clin. Pharmacol. Ther.* **106**, 810–820 (2019).
 23. Jonker, D.J. et al. Cetuximab for the treatment of colorectal cancer. *N. Engl. J. Med.* **357**, 2040–2048 (2007).
 24. Garrett, C.R. et al. Phase I dose-escalation study to determine the safety, pharmacokinetics and pharmacodynamics of brivanib alaninate in combination with full-dose cetuximab in patients with advanced gastrointestinal malignancies who have failed prior therapy. *Br. J. Cancer* **105**, 44–52 (2011).
 25. Van Cutsem, E. et al. Binimetinib, encorafenib, and cetuximab triplet therapy for patients with BRAF V600E-mutant metastatic colorectal cancer: safety lead-in results from the phase III beacon colorectal cancer study. *J. Clin. Oncol.* **37**, 1460–1469 (2019).
 26. Le Louedec, F. et al. Cetuximab pharmacokinetic/pharmacodynamics relationships in advanced head and neck carcinoma patients. *Br. J. Clin. Pharmacol.* **85**, 1357–1366 (2019).
 27. Azzopardi, N. et al. Cetuximab pharmacokinetics influences progression-free survival of metastatic colorectal cancer patients. *Clin. Cancer Res.* **17**, 6329–6337 (2011).
 28. Girard, P. et al. Drug-disease model of tumor size and overall survival in metastatic colorectal cancer patients treated with cetuximab administered weekly or every second week. European Cancer Congress, Amsterdam, Netherlands. (2013).
 29. Tan, A.R. et al. Pharmacokinetics of cetuximab after administration of escalating single dosing and weekly fixed dosing in patients with solid tumors. *Clin. Cancer Res.* **12**, 6517–6522 (2006).
 30. Taberero, J. et al. Cetuximab administered once every second week to patients with metastatic colorectal cancer: a two-part pharmacokinetic/pharmacodynamic phase I dose-escalation study. *Ann. Oncol.* **21**, 1537–1545 (2010).
 31. Bokemeyer, C. et al. Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. *J. Clin. Oncol.* **27**, 663–671 (2009).
 32. Van Cutsem, E. et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N. Engl. J. Med.* **360**, 1408–1417 (2009).
 33. Cheng, A.-L. et al. Efficacy, tolerability, and biomarker analyses of once-every-2-weeks cetuximab plus first-line FOLFOX or FOLFIRI in patients with KRAS or all RAS wild-type metastatic colorectal cancer: the phase 2 APEC study. *Clin. Colorectal Cancer* **16**, e73–e88 (2017).
 34. Fernandez-Plana, J. et al. Biweekly cetuximab in combination with FOLFOX-4 in the first-line treatment of wild-type KRAS metastatic colorectal cancer: final results of a phase II, open-label, clinical trial (OPTIMIX-ACROSS Study). *BMC Cancer* **14**, 865 (2014).
 35. Wicha, S.G., Chen, C., Clewe, O. & Simonsson, U.S.H. A general pharmacodynamic interaction model identifies perpetrators and victims in drug interactions. *Nat. Commun.* **8**, 2129 (2017).
 36. Genz, A. Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Stat.* **1**, 141–149 (1992).
 37. Soetaert, K., Petzoldt, T. & Setzer, R.W. Solving differential equations in R: package desolve. *J. Stat. Softw.* **33**, 9 (2010).
 38. Kuhn, E. & Lavielle, M. Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Stat. Data Anal.* **49**, 1020–1038 (2005).
 39. Ahn, J.E., Karlsson, M.O., Dunne, A. & Ludden, T.M. Likelihood based approaches to handling data below the quantification limit using NONMEM VI. *J. Pharmacokinet. Pharmacodyn.* **35**, 401–421 (2008).
 40. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
 41. Clopper, C.J. & Pearson, E.S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404–413 (1934).
 42. Viele, K. et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm. Stat.* **13**, 41–54 (2014).
 43. Sheiner, L.B. Is intent-to-treat analysis always (ever) enough? *Br. J. Clin. Pharmacol.* **54**, 203–211 (2002).
 44. Holford, N.H.G. & Anderson, B.J. Allometric size: The scientific theory and extension to normal fat mass. *Eur. J. Pharm. Sci.* **109S**, S59–S64 (2017).
 45. Sullivan, D.C., Schwartz, L.H. & Zhao, B. The imaging viewpoint: how imaging affects determination of progression-free survival. *Clin. Cancer Res.* **19**, 2621–2628 (2013).
 46. Mandrekar, S.J. et al. Evaluation of alternate categorical tumor metrics and cut points for response categorization using the RECIST 1.1 data warehouse. *J. Clin. Oncol.* **32**, 841–850 (2014).
 47. Paller, C.J. et al. Design of phase I combination trials: recommendations of the clinical trial design task force of the nci investigational drug steering committee. *Clin. Cancer Res.* **20**, 4210–4217 (2014).
 48. Foo, L.K. & Duffull, S. Adaptive optimal design for bridging studies with an application to population pharmacokinetic studies. *Pharm. Res.* **29**, 1530–1543 (2012).
 49. Pierrillas, P.B. et al. Model-based adaptive optimal design (MBAOD) improves combination dose finding designs: an example in oncology. *AAPS J.* **20**, 39 (2018).
 50. Mentré, F., Mallet, A. & Baccar, D. Optimal design in random-effects regression models. *Biometrika* **84**, 429–442 (1997).

© 2020 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of the American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

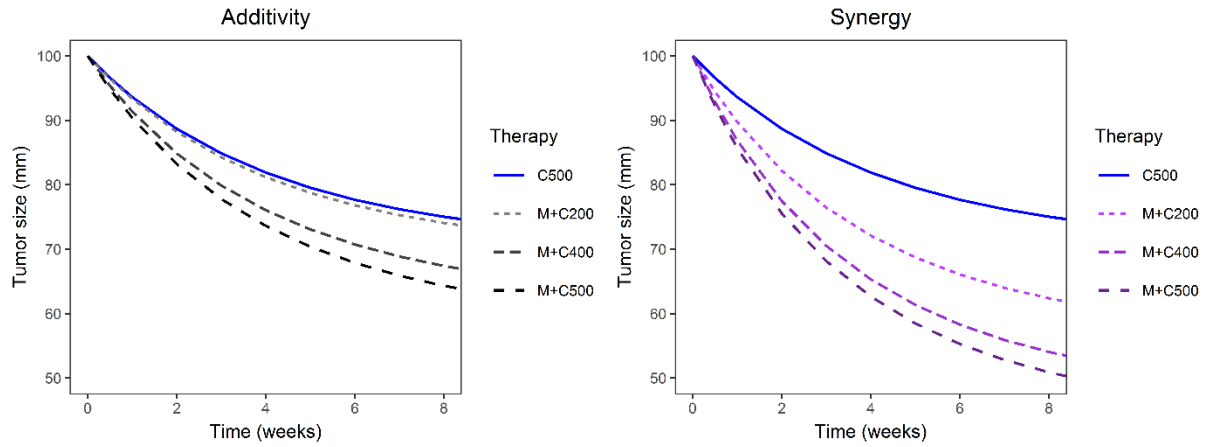


Figure S1. Median tumor size for the additivity (left) and synergy (right) scenario with different therapies. Cetuximab alone at 500 mg/m² every 2 weeks (C500) in blue and solid line, or combination treatment with different doses of cetuximab: 200, 400, or 500 mg/m² every 2 weeks (respectively M+C200, M+C400 and M+C500, respectively) in different shades of grey and purple for additivity and synergy, respectively.

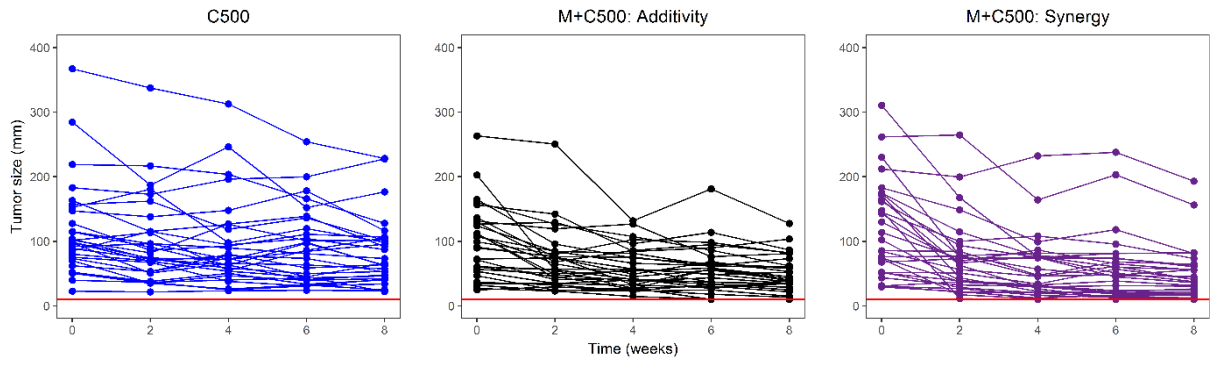
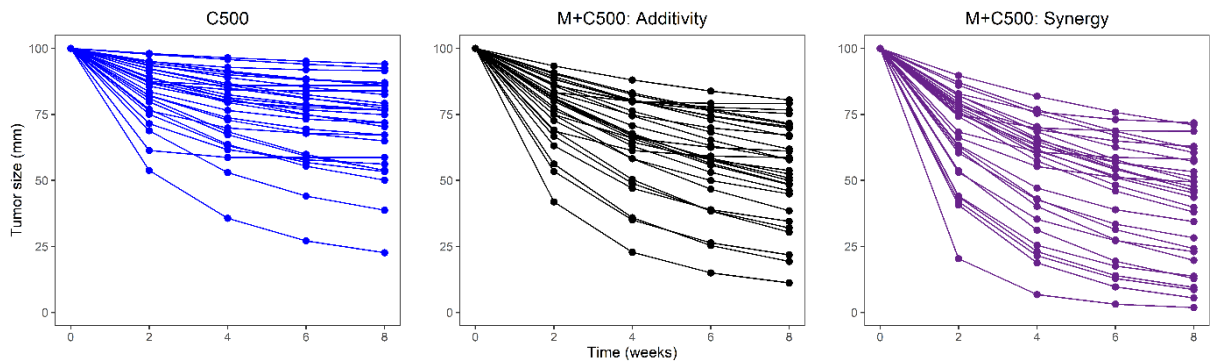


Figure S2. Simulated tumor size data for 30 individuals under cetuximab alone (C500, in blue) or combination treatment (M+C500) for the additivity (in black) or synergy scenario (in purple). (A) The red horizontal line represents the lower limit of quantification of 10 mm.



(B) Without residual error and an initial tumor size of 100 mm.

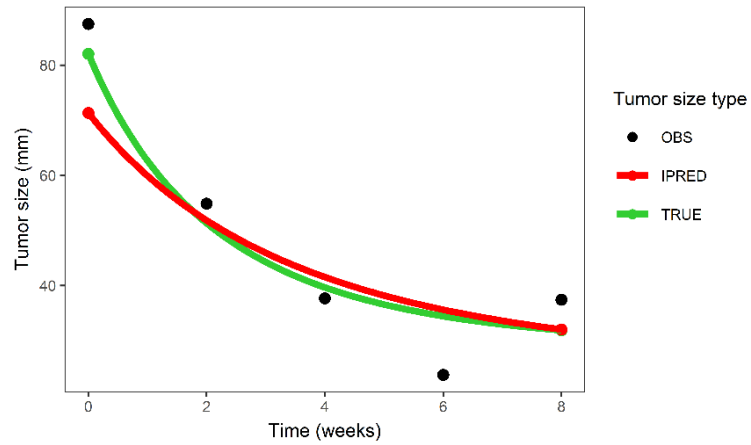


Figure S3. Simulated observations (OBS, black dots), model based predictions (IPRED, green line and dots), and true values (TRUE, red line and dots) for tumor sizes in one individual.

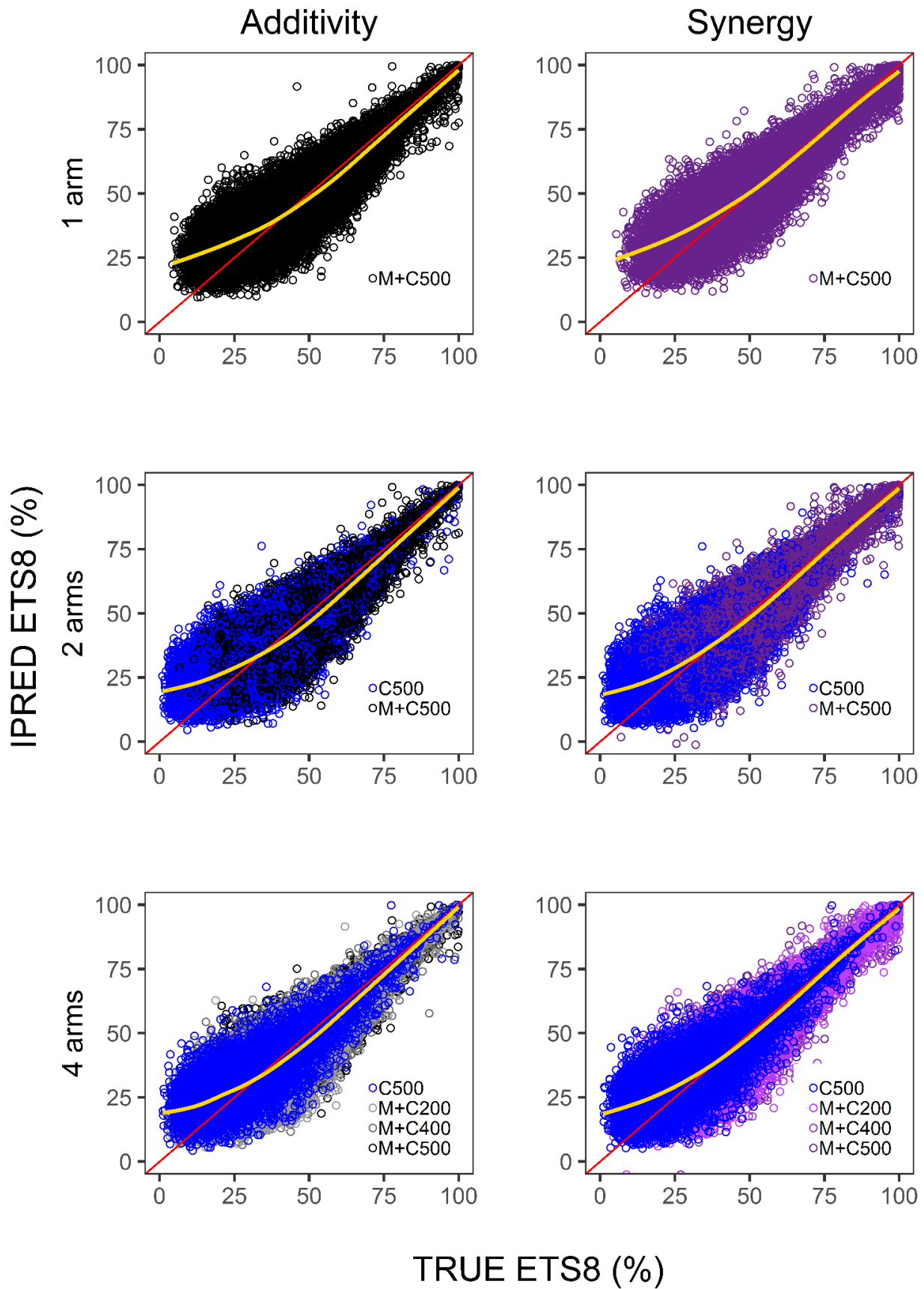


Figure S4. Individual model-based (IPRED) vs. true ETS8 (TRUE) values with different designs by row and under different interaction scenarios by column.

Troisième partie : Modélisation de l'interaction bactériophages/bactéries pour le traitement de la pneumonie

Dans le contexte de l'inquiétante augmentation de l'émergence et de la propagation des bactéries multi-résistantes, il est nécessaire de mener des actions contre la résistance aux antibiotiques avec une forte incitation à mettre au point de nouvelles approches thérapeutiques (O'Neill 2014). Parmi les solutions les plus prometteuses et radicalement différentes des molécules classiques, la phagothérapie suscite de plus en plus l'intérêt des communautés scientifiques et médicales (Caflisch et al. 2019). Cette approche utilise des bactériophages *i.e.* des virus qui n'infectent que les bactéries. Cette phagothérapie fût rapidement adoptée dans un premier temps, mais le manque de connaissance scientifique sur les bactériophages, ainsi que le rapide développement de la production industrielle des antibiotiques a sonné le glas de ce traitement, à l'exception de certains pays de l'Europe de l'Est où la phagothérapie s'est fortement développée. Les résultats prometteurs des études animales suggèrent que la phagothérapie pourrait être envisagée comme une alternative sérieuse aux antibiotiques. Néanmoins, il reste encore beaucoup à apprendre avant que l'on puisse proposer un traitement à base de phages. En particulier, il est nécessaire de comprendre et de quantifier les interactions phages/bactéries afin de définir les modes d'administration de phages les plus appropriés.

3.1. Résumé

Les travaux présentés s'inscrivent dans un projet global impliquant l'Institut Pasteur (équipe « Bactériophage, bactérie, hôte ») et IAME (équipes EVRest et BIPID) qui a pour but de mieux appréhender la phagothérapie. La série d'expériences présentée a été menée sur les bactériophages 536P1, dont les propriétés thérapeutiques ont été précédemment validées dans un modèle d'infection pulmonaire aiguë à *E.coli* 536 (Dufour et al. 2015). Nous avons construit un modèle mathématique à partir de ces données afin de caractériser les aspects essentiels concernant les interactions entre phages et bactéries. Au cours de cette nouvelle étude, des expériences *in vitro*, puis chez la souris, ont été conduites pour le couple 536/536_P1, et l'analyse successive de ces données nous a permis de construire un modèle mathématique décrivant l'évolution de l'infection bactérienne et l'efficacité de la phagothérapie.

Les données *in vitro* consistent en trois types d'expériences. Pour l'étude de *one-step growth*, la multiplication des phages est observée pour un unique cycle lytique. L'étude de co-évolution a permis de suivre les concentrations bactérienne et virale de manière longitudinale. Enfin, les cinétiques de lyses permettent de suivre la concentration bactérienne selon différentes doses connues de phages. Ces expériences ont notamment permis d'estimer les paramètres du cycle lytique du bactériophage 536_P1 en présence de la bactérie 536.

Les études chez la souris saine, où la concentration pulmonaire de phages a été mesurée lors du sacrifice (à différents temps selon les souris), ont permis de caractériser la pharmacocinétique des bactériophages, après administration intraveineuse (IV, N = 37) ou intratrachéale (IT, N = 16).

Enfin, des souris ont été infectées puis traitées ou non par bactériophages à différentes doses et voies d'administration (IV et IT). Grâce à une technique originale basée sur l'imagerie des bactéries bioluminescentes, les charges bactériennes pulmonaires ont pu être quantifiées à différents temps chez les mêmes animaux. La concentration pulmonaire de phages a été relevée uniquement au temps de sacrifice (si la souris survivait jusqu'au temps prévu).

Certains paramètres du modèle final décrivant les données de souris infectées (N = 183) ont été fixés à partir des estimations obtenues à partir de modèles construits en amont, notamment celui décrivant les données *in vitro* et celui décrivant la pharmacocinétique des phages chez la souris saine. L'ensemble des paramètres ont été estimés par MNLEM, avec l'algorithme SAEM implémenté dans le logiciel Monolix (version 2018R2), à l'exception des paramètres de la pharmacocinétique des phages estimés par *naïve pooling of data*. Le modèle conjoint final a été évalué par VPC corrigées (Bergstrand et al., 2011) de la charge bactérienne et par VPC de survie. A partir du modèle final, il a été étudié l'impact de la dose et de la voie d'administration des phages, de l'inoculum bactérien et des paramètres du cycle lytique sur la dynamique médiane des phages et des bactéries.

Les principaux paramètres du cycle lytique ont été déterminés à partir des expériences *in vitro* et il a notamment été estimé que 570 phages sont relargués lors de la lyse d'une bactérie 536 (*burst size*). A partir des données chez la souris non infectée, c'est-à-dire sans amplification possible des phages, l'administration de phages par voie IV conduisait à une exposition 5000 fois moindre qu'après administration locale (IT). L'analyse des données chez la souris infectée, afin de caractériser les paramètres clés de l'interaction phage-bactérie, a nécessité l'utilisation d'un système d'équations différentielles ordinaires combinant un modèle de cinétique bactérienne et un modèle de dynamique virale. Le modèle final consiste en un modèle proie-

prédateur (bactérie-bactériophage, respectivement) qui intègre le rôle du système immunitaire sur la décroissance bactérienne. Le modèle a montré que pour un faible inoculum bactérien, par exemple avec une concentration bactérienne initiale de 6 log CFUeq/g, le système immunitaire seul permet de contrôler l'infection. Pour un inoculum modéré (aux alentours de 7 log CFUeq/g), les phages et le système immunitaire agissent de concert pour parvenir à réduire la charge bactérienne. Enfin, lorsque l'inoculum bactérien est très important (par exemple supérieur à 8 log CFUeq/g), la phagothérapie ne permet pas de contrôler la charge bactérienne. Le modèle montre aussi un lien significatif ($p < 10^{-3}$ par test de Wald) entre la charge bactérienne des souris et leur survie, par le biais d'un modèle conjoint. Le risque instantané de décès était multiplié par 18 pour une charge bactérienne augmentée de 2 log CFUeq/g, ce qui correspond à un *hazard ratio* de 4,2. Ce modèle était correctement évalué par les différents graphiques diagnostiques et a permis de décrire les différents profils individuels malgré une forte variabilité entre ces derniers.

Le modèle final a permis de quantifier l'efficacité de la phagothérapie en fonction de la dose et de la voie d'administration : pour un inoculum bactérien létal (7 log CFUeq/g) pour 100 % des souris non traitées, l'administration de phages, permettait un contrôle de l'infection bactérienne chez la plupart des individus, pour tous les groupes de doses et de voie d'administration. En médiane, la charge bactérienne commençait à décroître moins d'une demi-heure après administration IT à haute dose (IT Hi), environ une heure et demie après IT à faible dose (IT Lo, dose 1000 fois plus importante que la dose faible), et moins de 4 h après administration IV à faible dose (IV Lo). Ces différences, notamment entre les voies d'administrations, pouvaient devenir plus importantes pour un phage présentant un *burst size* 10 fois moins important. Les charges bactériennes commençaient en médiane à décroître à 0,3, 2,7 et 11,3 h après administration, chez les IT Hi, IT Lo et IV Lo, respectivement. Ces travaux montrent que la voie locale est à privilégier par rapport à la voie IV, en particulier lorsque les phages présentent une mauvaise capacité d'amplification. Le modèle construit permet donc bien d'évaluer l'efficacité de la phagothérapie en fonction de la dose et du mode d'administration. Ce travail permettra d'aider au choix de protocole lors de futures études sur les bactériophages. Un article présentant ce travail est soumis.

3.2. Article 4 (soumis)

*Combination of in vivo phage therapy data with in silico model highlights key parameters for
treatment efficacy*, soumis pour publication

Raphaëlle Delattre*, Jérémy Seurat*, Feyrouz Haddad, Thu Thuy Nguyen, Baptiste Gaborieau,
Rokhaya Kane, Nicolas Dufour, Jean-Damien Ricard, Jérémie Guedj, Laurent Debarbieux

Title

Combination of *in vivo* phage therapy data with *in silico* model highlights key parameters for treatment efficacy

Authors

Raphaëlle Delattre^{1,2,3*}, Jérémy Seurat^{2*}, Feyrouz Haddad^{1,2}, Thu-Thuy Nguyen², Baptiste Gaborieau^{1,2,3}, Rokhaya Kane¹, Nicolas Dufour⁴, Jean-Damien Ricard², Jérémie Guedj^{2**} et Laurent Debarbieux^{1**}

Affiliations

1 Institut Pasteur, Bacteriophage, Bacterium, Host Laboratory, Paris F-75015 France

2 Université de Paris, INSERM, IAME, Paris F-75006, France

3 Université de Paris, Cellule Pasteur, Paris F-75013, France

4 Centre Hospitalier René Dubos, Cergy Pontoise F-95503, France

* : co-first authors

** : co-corresponding authors

Lead contact: Laurent Debarbieux (laurent.debarbieux@pasteur.fr)

Summary

The clinical (re)development of phage therapy to treat antibiotic resistant infections requires grasping specific biological properties of bacteriophages (phages) as antibacterial. However, identification of optimal dosing regimens is hampered by the poor understanding of phage-bacteria interactions *in vivo*. Here we developed a general strategy coupling *in vitro* and *in vivo* experiments with a mathematical model to characterize the interplay between phage and bacterial dynamics during pneumonia induced by a pathogenic strain of *Escherichia coli*. The model estimates some key parameters for phage therapeutic efficacy, in particular the impact of dose and route of administration on phage dynamics and the synergism of phage and the innate immune response on the bacterial clearance rate. Simulations predict a low impact of the intrinsic phage characteristics in agreement with the current semi-empirical choices of phages for compassionate treatments. Model-based approaches will foster the deployment of future phage therapy clinical trials.

Keywords

Bacteriophage, antibiotic resistance, bacterial infection, pneumonia, mathematical model, innate immunity, simulation

Introduction

The pronounced slowdown in the discovery of new antibiotics to treat bacterial infections caused by multidrug resistant (MDR) pathogens has revived the interest for bacteriophages (phages), viruses infecting bacteria (Gordillo Altamirano and Barr, 2019; Kortright et al., 2019). Pre-clinical experiments and several clinical cases reports have demonstrated the potential of phages to successfully treat infections caused by MDR bacteria (Aslam et al., 2020; Corbellino et al., 2020; Jennes et al., 2017; Melo et al., 2020; Schooley et al., 2017). These success are nonetheless balanced by clinical studies that did not confirm these positive results (Jault et al., 2019; Leitner et al., 2020; Sarker et al., 2016), showing that translation from experimental results and promising individual treatments to clinical trials remains challenging.

A major milestone for a widespread use in human populations is the determination of the optimal dose, route of administration and treatment duration. This is a particularly complex endeavor for phages, as standard assessments of clinical pharmacology used to determine the processes of Administration, Distribution, Metabolism and Excretion (ADME) of drugs are not adapted to phages. Indeed, phages massively replicate on their target bacteria with kinetic parameters governing this viral amplification (receptor recognition, genome amplification, particles assembly and bacterial lysis) being specific to each phage-bacteria pair (Dion et al., 2020). Furthermore, phages are naturally present in human-associated microbiotas and their rate and route of elimination do not follow usual metabolic pathways through kidney or liver, as for xenobiotic products (Dabrowska and Abedon, 2019).

Here we coupled *in vitro* and *in vivo* experiments with mathematical modeling to approach the key parameters of phage-bacteria-host interactions in order to guide treatment strategies. This approach has been successful in drug development for various chronic (HIV, HCV) and acute (Ebola, Zika, Influenza) viral infections (Friberg and Guedj, 2020; Perelson and Guedj, 2015). Unlike previous models that focused on specific aspects of phage-bacteria, phage-host immune system or bacteria-host immune system interactions (Cairns et al., 2009; Roach et al., 2017; Smith et al., 2011; Weld et al., 2004), we aimed to design a synthetic, semi-mechanistic model that could be used to understand the tripartite phage-bacteria-host interactions *in vivo*. We purposely designed a series of specific experiments (Figure 1) to estimate the central parameters from a previously validated experimental mice model of pulmonary infection using the bioluminescent *Escherichia coli* pathogenic strain 536 and the virulent phage 536_P1 (Dufour et al., 2015). Dynamic and direct microbiological data were collected from groups of animals that were either (bacterial) uninfected and (phage) treated, infected and untreated, or infected and treated, using different inoculum doses and routes of administration. These data led us to

develop an original mathematical model recapitulating the tripartite interactions over time. This model identified key parameters affecting bacterial kinetics *in vivo*. It was then exploited to test scenarios including variations of phage intrinsic properties. The use of such model will rationalize the phage choice, the dose and route of administration in order to optimize the efficacy of phage therapy.

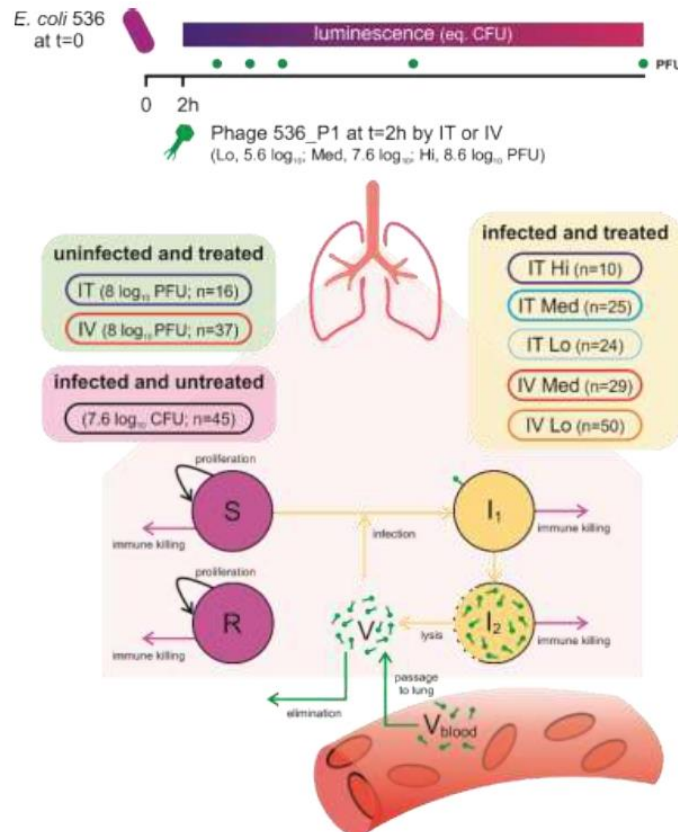


Figure 1. Schematic representation of the experimental design of the study.

In vivo experiments were conducted in mice that were either i) uninfected and treated (green box, with phages administered either intratracheally, IT, or intravenously, IV); ii) infected and untreated (magenta box); iii) infected and IT- or IV-treated at various doses (yellow box). The intensity of light emitted by *E. coli* strain 536-lux was recorded along all experiments and translated to bacterial lung density, while phage 536_P1 density was measured only once (upon sacrifice) in each animal at specific time points. A mathematical compartmental model was developed to fit to these data, with S, R, I₁ and I₂ representing the susceptible, refractory, infected non-productive, productively infected bacteria, respectively, and V and V_{blood} the phage concentration in lungs and blood, respectively.

Results

Biodistribution of phage in uninfected mice

We first studied the biodistribution of phage 536_P1 in uninfected mice following a single administration of $8 \log_{10}$ PFU by intratracheal (IT) or intravenous (IV) route (Figure 2). After IT administration, phage concentrations in the lungs varied over time only within 1-log with observed mean values between 7.0 and $8.0 \log_{10}$ PFU/g. By contrast, after IV administration, the observed peak concentration was reached four hours post-administration ($5.1 [4.7, 5.5] \log_{10}$ PFU/g), which was over 2-log below the observed concentration after IT administration ($7.2 \pm [6.8, 7.8] \log_{10}$ PFU/g). Therefore, after IV administration only a fraction F of the phage, estimated to 0.02%, ultimately reached the lung compartment, leading to exposure approximately 5000 times lower than after IT administration and an estimation of the absorption rate into the lung compartment, k_a , of 0.3 h^{-1} . (Eq. 1-2). The kinetics of phage elimination was also different with an estimated half-life of 12 and 3 hours after IT and IV administration, respectively (see $k_e(\text{IT})$ and $k_e(\text{IV})$ values in Table 1).

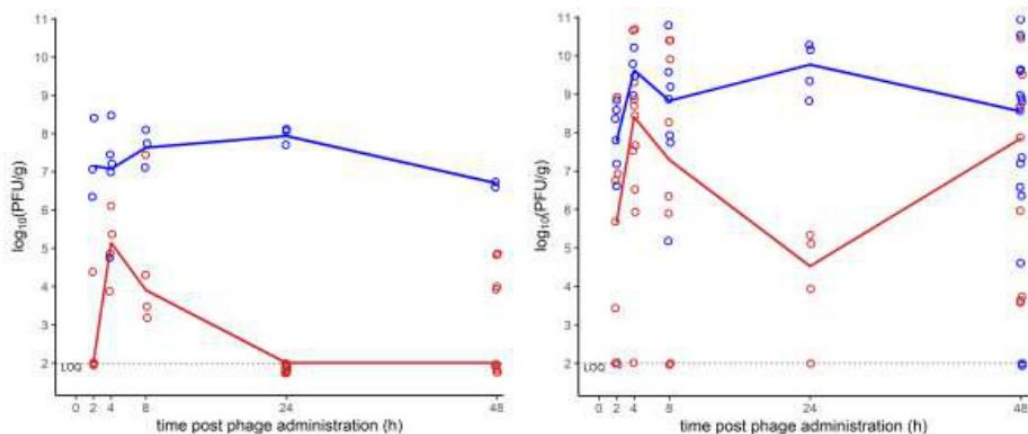


Figure 2. Distribution of phage 536_P1 in lungs of treated uninfected and treated infected mice.

Lungs of uninfected (left, $N=53$) and *E. coli* infected (right, $N=138$) mice were homogenized at 2, 4, 8, 24 or 48h post-administration of a single dose of phage 536_P1 ($8 \log_{10}$ PFU in uninfected mice and 5.6, 7.6, or 8.6 \log_{10} PFU in infected mice), either by intratracheal (blue) or intravenous (red) route. Lines represent the median phage concentrations at each time point. The dashed line represents the limit of quantification ($\text{LOQ}=2 \log_{10}$ PFU/g).

Table 1. Parameter estimates of the final phage/bacteria dynamics model in mice

Parameter	Name	Unit	Fixed/Estimated from	Fixed effect (rse, %)	Sd of random effect (rse, %)
Bacterial growth rate	α	h^{-1}	Untreated infected	0.21 (-)	0 (-)
Initial number of bacteria	B_0	$\log_{10} \text{CFU g}^{-1}$	Untreated infected	7.1 (1)	0.14 (6)
Bacterial load plateau	B_{max}	$\log_{10} \text{CFU g}^{-1}$	Untreated infected	10.4 (-)	0 (-)
Time for immunity activation	t_I	h	Untreated infected	3.0 (-)	0 (-)
Maximal immunity effect	δ_M	$\text{CFU g}^{-1} \text{h}^{-1}$	Untreated infected	$6.7 \cdot 10^6$ (-)	0.05 (-)
Bacterial load for 50% of δ	B_{50}	CFU g^{-1}	Untreated infected	$2.3 \cdot 10^7$ (-)	0 (-)
Infectivity rate	β	$\log_{10} \text{g CFU}^{-1} \text{h}^{-1}$	Treated infected	6.9 (7)	0.11 (43)
Transition rate	e	h^{-1}	<i>in vitro</i>	1.2 (-)	0 (-)
Lysis rate	δ_V	h^{-1}	Treated infected	0.09 (25)	0 (-)
Burst size	b	PFU CFU^{-1}	<i>in vitro</i>	570 (-)	0 (-)
Relative bioavailability (IV/IT)	F	-	Treated uninfected	0.0002 (-)	0 (-)
Phage absorption rate	k_a	h^{-1}	Treated uninfected	3 (-)	0 (-)
Phage elimination rate after IT	$k_e(IT)$	h^{-1}	Treated uninfected	0.058 (-)	0 (-)
Phage elimination rate after IV	$k_e(IV)$	h^{-1}	Treated uninfected	0.23 (-)	0 (-)
Phage distribution weight	W	g	Treated uninfected	1 (-)	0 (-)
Initial refractory proportion	f	-	Treated infected	0.105 (45)	2.8 (19)
Weibull shape	κ	-	Treated infected	1.8 (11)	0 (-)
Weibull scale	λ	h^{-1}	Treated infected	$2.4 \cdot 10^4$ (98)	0 (-)
Impact of bacterial kinetics on death	γ	g CFU^{-1}	Treated infected	1.44 (13)	0 (-)
Additive error on bacterial load	σ_B	$\log_{10} \text{CFU g}^{-1}$	Treated infected	-	0.47 (4)
Additive error on phage concentration	σ_V	$\log_{10} \text{PFU g}^{-1}$	Treated infected	-	2.6 (11)

Sd: standard deviation, rse: relative standard error

Bacterial kinetics in infected-untreated mice

IT administration of the *E. coli* strain 536 led to a large variability in the lung bacterial load of untreated animals, as measured by emitted luminescence and CFUeq/g (see methods), with observed values ranging between 5.0 and 9.9 \log_{10} CFUeq/g at two hours post infection (Figure 3). This heterogeneity in the effective bacterial inoculum mimics the diversity of bacterial loads in human patients with pneumonia (Jiang et al., 2014), and allowed us to consider each mouse as an individual when developing the mathematical model (see below and Eq. 3). This heterogeneity also informed on the threshold of bacterial density above which the immune response of the host cannot control the infection. Indeed, when the bacterial load was greater

than $6.8 \log_{10}$ CFUeq/g at two hours post-infection, the bacterial density increased continuously, while, below this threshold, we observed a rapid reduction of the bacterial density caused by the immune response to the infection (Figure 3).

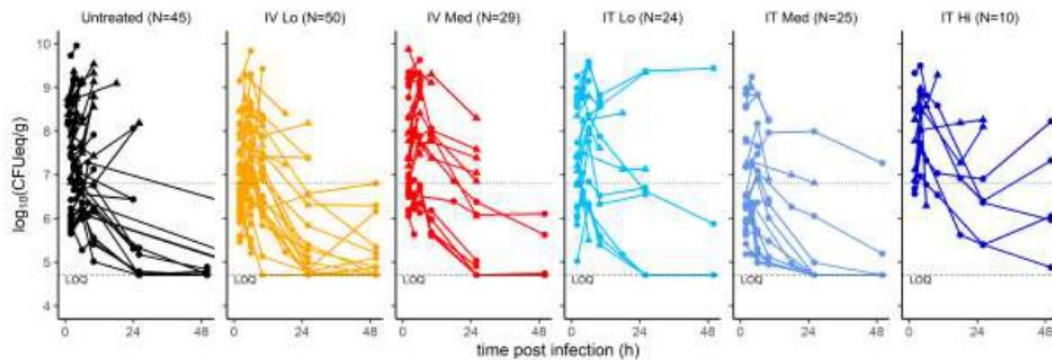


Figure 3. Kinetics of bacterial densities in untreated and phage-treated animals.

Bioluminescence data from infected animals were stratified by treatment groups: untreated (black), treated with phage 536_P1 administered by IV at a low dose ($5.6 \log_{10}$ PFU, yellow) or medium dose ($7.6 \log_{10}$ PFU, red), or by IT at either low dose (cyan), or medium dose (blue) or high dose ($8.6 \log_{10}$ PFU, dark blue). The dotted horizontal line at $6.8 \log_{10}$ CFUeq/g corresponds to the threshold value of bacterial load at two hours post-infection below which all animals controlled the infection. The dashed line represents the limit of quantification (LOQ= $4.7 \log_{10}$ CFUeq/g). Circles correspond to live animals until the time of sacrifice while triangles correspond to dead animals before the scheduled time of sacrifice.

Bacterial and phage kinetics in infected treated mice

Next, using the same infection procedure as above, mice (N=138) were treated two hours post-infection with various doses of phage 536_P1 (low dose: 5.6; medium dose: 6.6; high dose: 7.6 \log_{10} PFU), administered either by IT or IV route.

Phages rapidly replicated and their density in the lungs reached values 2 to 3- \log_{10} higher than in uninfected animals, showing that they encountered bacteria and efficiently replicated (Figure 2). At four hours post-phage administration (six hours post-bacterial infection), the median titer of phage in the lungs of IT-treated mice (N=59) was $9.6 [9.3, 9.9] \log_{10}$ PFU/g, as compared to $7.1 [7.1, 7.4] \log_{10}$ PFU/g in uninfected IT-treated mice (N=16). A similar difference was also observed in IV-treated animals (N=79), with phages concentration of $8.4 [7.0, 9.1] \log_{10}$ PFU/g as compared to $5.1 [4.7, 5.5] \log_{10}$ PFU/g in uninfected IV-treated animals (N=37). A dose

effect was noticed in IV-treated animals, with a median concentration of 7.6 [6.4, 8.5] \log_{10} PFU/g in the lungs of mice receiving a low dose (N=50), as compared to 10.6 [10.0,10.6] \log_{10} PFU/g in mice treated with a medium dose (N=29).

The initial effective median bacterial inoculum was similar in each of the three groups with levels of 6.9 [6.1, 8.0], 7.2 [6.3, 8.5] and 6.7 [6.3, 7.8] \log_{10} CFUeq/g in the lungs of untreated, IT- and IV-treated and animals, respectively (Figure 3). Afterwards, two hours post-phage administration, the bacterial loads tended to be lower in treated (N=138) compared to untreated (N=45) animals, with levels of 7.5 [6.6, 8.8] \log_{10} CFUeq/g and 8.4 [7.6, 8.8] \log_{10} CFUeq/g, respectively ($p=0.17$, Wilcoxon-Mann-Whitney test). However, at this time point there was no effect of the route of administration, with values of 7.5 [6.5, 8.8] \log_{10} CFUeq/g and 7.2 [6.6, 8.5] \log_{10} CFUeq/g in IT and IV-treated groups, respectively ($p=0.27$ and $p=0.14$ to untreated, respectively, and $p=0.75$ when comparing IT and IV-treated groups, Wilcoxon-Mann-Whitney test).

The mortality rates at 48 hours were similar in all groups (respectively 42%, 48% and 52% in IT, IV and untreated animals) (Figure 4). Mortality rate was not dependent on the dose of administered phages (at 48 hours equal to 51%, 25% and 50% in animals receiving a low, medium or high dose of phages, respectively; $p=0.92$, 0.14 and 0.68 to untreated, logrank test), (Figure S4). However, it is important to note that in the subgroup of animals with a high effective bacterial inoculum ($>6.8 \log_{10}$ CFUeq/g), treatment tended to increase the survival rate (Figure 4, $p=0.07$ in IT and $p=0.21$ in IV, as compared to untreated, respectively, logrank test), whereas all animals with a low effective inoculum ($<6.8 \log_{10}$ CFUeq/g) survived at 48h, regardless of treatment (Figure 4).

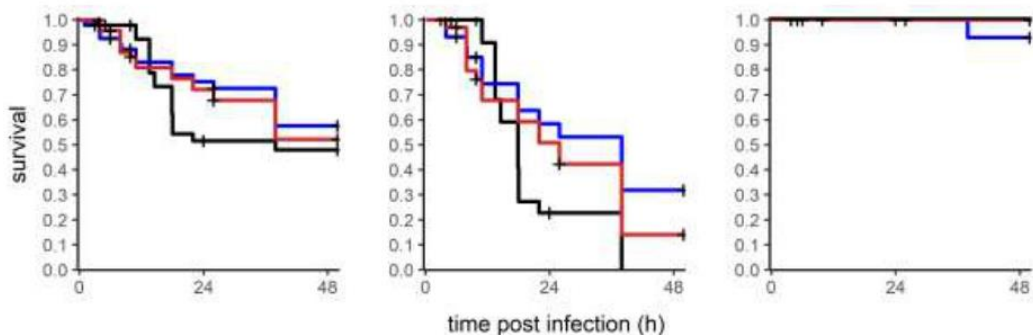


Figure 4. Survival of infected and treated animals in function of the pulmonary bacterial load at two hours post-infection

The survival curves were obtained from animals that were infected untreated (black, N=45), infected treated by IT route (blue, N=59) or infected treated by IV route (red, N=79). The left panel displays all infected animals (N=183). The middle panel includes only animals (N=78) with a pulmonary bacterial load at two hours post-infection greater than $6.8 \log_{10}$ CFUeq/g. The right panel includes only animals (N=99) with a pulmonary bacterial load at two hours post-infection lower than $6.8 \log_{10}$ CFUeq/g. Crosses represent sacrifice time. Of note, 6 animals sacrificed at two hours post infection were not represented in both middle and right panel.

Unraveling bacteria-phage interaction using a mathematical model

Next, we used these data in both infected and uninfected animals together with parameters determined *in vitro* (see methods) to build a mathematical model of phage-bacteria-host interaction (Eq. 4-10, Figure 1). Fitting the model to bacterial load data, we first determined that a model accounting for the control of bacterial growth by the innate immune response was needed to recapitulate the threshold effect associated with the effective bacterial inoculum. This phenomenon could be well captured using our model (Eq. 3), with estimates of δ_M , the maximal decay induced by the immune response, of $6.8 \log_{10}$ CFUeq/g/h and B_{50} , the bacterial load needed to reach $\delta_M/2$, of $7.4 \log_{10}$ CFUeq/g. This corresponds to a threshold for bacterial control equal to $6.94 \log_{10}$ CFUeq/g. Thus, if the administered bacterial inoculum, B_0 , was below this value, bacteria were rapidly eliminated by the immune system with a rate $\sim \delta_M/B_{50} = 0.3 \text{ h}^{-1}$, corresponding to a half-life of 2.3 hours. However, when the bacterial inoculum was above this value, the model predicts a saturation of the immune response, enabling bacterial growth with a bacterium doubling time estimated at 3.5 hours.

We performed *in vitro* experiments from which we estimated two key parameters of phage dynamics that could not be identified *in vivo*, namely the median eclipse phase and the burst size (see methods and Supplementary Figure S2). Assuming that these parameters had the same value *in vivo*, we could estimate the infection rate of bacteria by phages, noted β , to $6.9 \log_{10}$ g/CFU/h, and the elimination rate of productively infected bacteria, δ_V , to 0.09 h^{-1} (eg, a half-life of 7.7 h).

Finally, we also used the model to assess the impact of bacterial kinetics on survival. Assuming a log-linear effect of bacterial load on the instantaneous risk of death, $h(t)$, the bacterial load was found to be highly associated with death ($\gamma=1.44$, $p<10^{-3}$, Wald test). This corresponds to a hazard ratio of 4.2; in other words, an increase of 2 log of the bacterial load raised the instantaneous risk of death by ~ 18 . The joint model developed could simultaneously

recapitulate bacterial kinetics, phage counts and survival rate (Supplementary Figures S5-S6), despite the heterogeneity of the individual profiles experimentally observed (Figure 5).

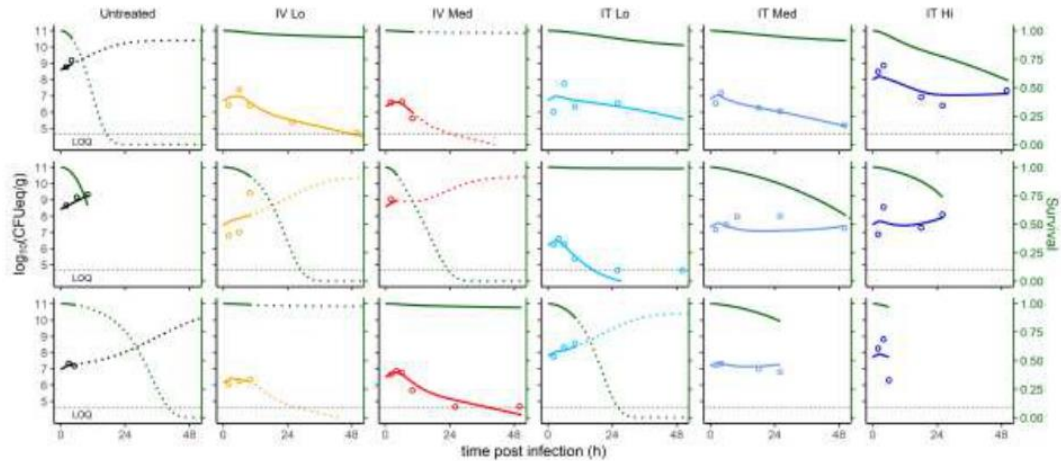


Figure 5. The mathematical model recapitulates the individual variations of the pulmonary bacterial loads and estimates the survival of infected animals.

Data and model predictions for three mice (lines) randomly chosen from each experimental groups (columns) are shown. The circles correspond to the pulmonary bacterial loads, the color lines correspond to the individual predictions from the model, and the green line represents the predicted survival probability. Lines are solid between $t=0$ and the last observation time point (death or pre-specified sacrifice) and are dotted after the pre-specified sacrifice. Interrupted lines indicate the death of animals before the pre-specified time point. The horizontal dashed line represents the limit of quantification ($LOQ=4.7 \log_{10} \text{CFUeq/g}$).

Effects of bacterial inoculum, dose, route of administration and phage characteristics on the bacterial kinetics

The model was then used to identify the key factors governing bacterial kinetics. Unsurprisingly, the bacterial inoculum, B_0 , was the most important parameter. In animals with $B_0 < 6 \log_{10} \text{CFUeq/g}$, bacterial loads rapidly decreased below the limit of quantification within 48h, regardless of treatment. In animals with large bacterial inoculum $B_0 > 8 \log_{10} \text{CFUeq/g}$, phage therapy was not sufficient to control the continuous growth of bacteria, regardless of the route and administered dose (Figure 5, Figure 6). The benefit of phage therapy was visible mostly for animals having an intermediate effective inoculum size, e.g. $6 < B_0 < 8 \log_{10} \text{CFUeq/g}$, in which phages were critical to prevent bacteria growth and to maintain the bacterial load under

control by the immune system. In those animals, the decline of bacterial load was biphasic, with a rapid decline in the first four hours of treatment, followed by a slower decline afterwards. This slower second phase was attributed in our model to the unrestricted replication of a sub-population of bacteria that were not eliminated by phages, representing both phage-resistant and phage-inaccessible (spatially out of reach) bacteria. The initial proportion of this phage-refractory population was estimated to 10% (Table 1), but could be variable across individuals (from 0.03% to 76%). The dynamics of refractory bacteria was then individual dependent, leading either to a slower but continuous bacterial decline or to a rebound of fully refractory bacteria within 24-48 hours (see individual examples in Figure 5). Importantly, the need to incorporate a compartment of refractory bacteria in the model was strongly supported by data and led to a significant improvement of the data fitting ($p < 10^{-3}$, Likelihood Ratio Test, not shown).

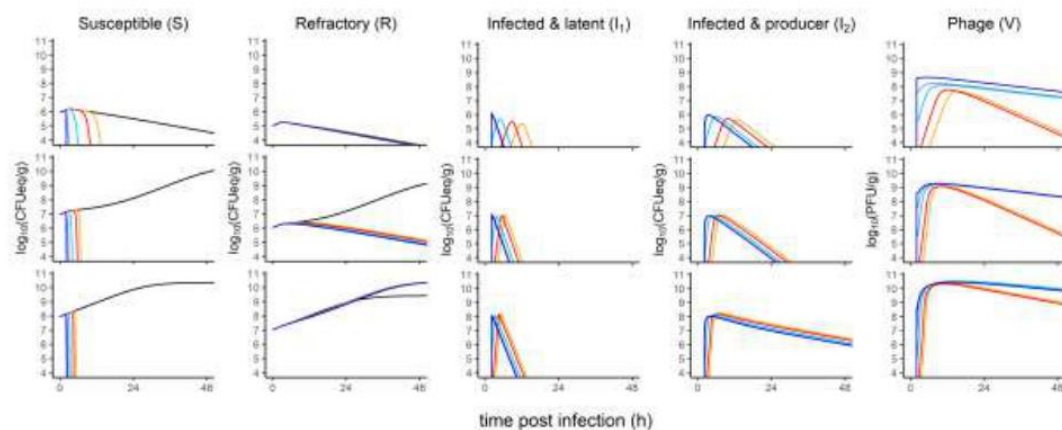


Figure 6. Simulations of different bacterial inoculums inform on the impact of refractory bacteria

Median profile assuming initial bacterial inoculums of 6, 7 and 8 \log_{10} CFUeq/g (top, middle and bottom lines, respectively). Predictions were calculated for each bacterial population (S, R, I_1 and I_2) and phages (V). The colors are indicative of both the route of administration (IT or IV) and the dose of phage received (Lo: 5.6 \log_{10} PFU; Med: 7.6 \log_{10} PFU; Hi: 8.6 \log_{10} PFU): untreated (black), IV-Lo (orange), IV-Med (red), IT-Lo (cyan), IT-med (blue), IT-high (dark blue).

The raw data did not allow to observe a role of route of administration in treatment outcome, except in the case of a low bacterial inoculum ($B_0=6 \log_{10}$ CFUeq/g) in which the rapid clearance of bacteria limited the encounter rate with phage. Our simulations indicated that IT and IV otherwise reached similarly high phages concentrations in lungs (Figure 6). IT administration was however associated with a substantially faster time to peak viral load, and this difference was particularly obvious with a low or an intermediate bacterial inoculum. For instance, with an intermediate bacterial inoculum ($B_0=7 \log_{10}$ CFUeq/g), the time to peak viral load varied between 1.6 h (IT high dose) and 6.0 h (IV low dose). As a consequence of this delay in phage maximal replication, IV route had a less efficiency to stop bacterial growth, with a time to peak bacterial load that could be as large as 4.0 h (IV low dose), to less than 1.1 h with IT route, and as low as 0.3 h in case of IT high dose.

Finally, we also used the model to explore the effect of intrinsic phage characteristics, namely burst size and lysis rate, on the bacterial kinetics. A 10-fold reduction in burst size (*i.e.* 57 PFU/CFU) predicted a larger impact of the route of administration, with a time to bacterial peak of 10 h post-IV compared to less than 3 h post-IT administration, even when using a low dose (Figure S7). However, a higher lysis rate value counter-balanced the effects of a low burst size value (Figure S8). These showed the sensitivity of the model to phage characteristics.

Discussion

Despite many encouraging data from both ancient literature and recent compassionate treatments, phage therapy remains underexploited, notably because of ambiguous results from clinical trials (Luong et al., 2020). Indeed, the translation from *in vitro* to clinical conditions remains unmet, in particular because using phages as a drug requires a specific method to elaborate a rational scheme for safe and optimal administration. While some studies performed *in silico* or in laboratory conditions (Cairns et al., 2009; Kasman et al., 2002; Leung and Weitz, 2017; Payne and Jansen, 2001, 2003) led to propose several mathematical models of phage-bacteria interactions, very few studies including the host immune response were conducted *in vivo* (Dabrowska and Abedon, 2019; Roach et al., 2017). Here, we developed a pharmacometric model, exploiting *in vivo* data from a murine pneumonia treated by a single virulent phage administered at three doses and via two administration routes, which robustness could support further the (re)deployment of phage therapy.

Our experimental approach was inspired from the work performed during the development of new drugs, where intensive pharmacokinetic and pharmacodynamic data are required to optimize the dosing regimens of a treatment (Marshall et al., 2019). Here, we collected an unprecedented biological set of data using an *in vivo* imaging tool to longitudinally follow the same animals over time and from which we developed a semi-mechanistic model to characterize the dynamics of a phage treatment.

The biodistribution of phages in the lungs of uninfected mice showed that phage loads were much higher and their exposition lasted longer following local (IT) compared to systemic (IV) administration. Consequently, IT administration led to a rapid decline of bacterial loads within the next two hours post-administration. Nevertheless, 12 hours post local and systemic phage administrations, all dosing regimens led to comparable levels of phage exposure, due to the large burst size of phage 536_P1. Despite being in agreement with previous literature (Morello et al., 2011; Oechslin et al., 2017), these biodistribution data cannot be yet generalized to all phages given the very large diversity of these viruses. Also, we did not study repeated IV administered doses that could allow for longer and higher phage exposition in lungs.

The heterogeneity of the initial bacterial loads observed two hours post-infection explains the lack of statistical differences in Kaplan-Meier survival estimations when pooling mice to compare routes of administration and doses. The source of this heterogeneity was the

consequence of choices made to maximize the probability of interactions between phages and bacteria, compared to our previous work using the same couple of phage-bacteria (Dufour et al., 2015). Here, we increased the inoculum and we used the direct intratracheal route (instead of the intranasal route) to reach the lung, in order to reduce spatial dispersion along the upper respiratory tract. The large range of situations observed was instrumental to enrich the mathematical model with multiple events (N=236 analyzed mice) to improve the assessments of phage-bacteria parameters, doses and routes of administration. Moreover, a direct consequence of the above conditions was the precise quantification of the threshold value ($6.94 \log_{10}$ CFUeq/g) above which the immune response was overwhelmed, which was very close to the value ($6.80 \log_{10}$ CFUeq/g) deduced from our model.

The modeling of phage-bacteria interactions in the host required the definition of a refractory bacterial population. This compartment could not be phenotypically characterized but encompasses bacteria that are escaping phage predation because they are phage-resistant and/or they remain spatially inaccessible to phages (Leung and Weitz, 2017). The inclusion of this compartment is supported by both the emergence of phage resistance observed *in vitro* after 10 h of incubation with phage 536_P1 (Supplementary Figure S2), and previous experimental and theoretical studies showing that the spatial distribution of bacteria in organs can prevent bacteria to be infected by phages (Lourenco et al., 2020; Sousa and Rocha, 2019). This was also in agreement with the modelling of phage therapy in immunocompromised hosts (Roach et al., 2017). Despite the large variation of the bacterial loads that directly affects the initial density of this refractory population, our model could appropriately recapitulate the kinetics of phage-bacteria interactions for each individual mouse, including the survival probability, testifying of its robustness. In addition, simulations performed with larger or lower burst sizes and lysis rates, showed that the model predicts outputs that are in agreement with results from animal studies in which different phages were used to treat bacterial infections (Alemayehu et al., 2012; Carmody et al., 2010; Henry et al., 2013; Morello et al., 2011).

To ensure the model recapitulates kinetics of phage-bacteria interactions during therapy, it was necessary to use as much as possible parameters deduced from animal experiments instead of *in vitro* conditions. For instance, growth rate of bacteria, infectivity rate and lysis rate, were all deduced from experimental data. Nevertheless, other parameters should be more deeply investigated to improve the fidelity of the model. For instance, growth of refractory population was approximated as similar to the susceptible bacteria, neglecting the possibility that the

fitness of mutants might be affected in the host. Indeed, such mutants could be, either or both, less virulent and strongly affected in their growth, allowing immune cells to eliminate them more effectively (Oechslin et al., 2017; Pouillot et al., 2012; Smith et al., 1987). The involvement of the immune response in the success of respiratory phage therapy was previously demonstrated and modelled, but remains mechanistically poorly explored (Roach et al., 2017). In this work we chose to account for the immune response in a synthetic way which encompasses mechanical factors, such as mucins and surfactant proteins, as well as various cellular actors of the innate immune system (Bergeron et al., 1998; Marriott and Dockrell, 2007).

Predictions from the mathematical model showed the importance of viral load, burst size and lysis rate to fully capture bacterial kinetics and animal outcome. Therefore, in a clinical context, the administration route, and to a lesser degree the dose, will play a critical role in the rate at which bacterial growth will be controlled, supporting the determination of key parameters of phage-bacteria interactions as a prerequisite for optimizing phage therapy. However, the model predicted that substantial variations of the intrinsic characteristics of phages (burst size and lysis rate) were not jeopardizing treatment output, which fits with the current semi-empirical choice of phages for compassionate treatments. Indeed, currently phages are chosen exclusively for their capacity to infect bacteria on *in vitro* conditions where phage-bacteria interactions are artificially optimized. The model is therefore robust, which argues in favor of its relevance for many other phages.

To conclude, the model developed during this work could be used to predict the efficacy of virtually any phage for which a minimum set of *in vitro* and *in vivo* data must be obtained, which will considerably lower the number of experiments needed to validate such a phage during pre-clinical development. Beyond phage therapy, the model could also be implemented to test combined anti-infectious therapies such as the association of phages with antibiotics (Lin et al., 2020).

Methods

Determination of *in vitro* infection parameters of phage 536_P1

Parameters of the interactions between strain 536 and phage 536_P1 have been determined *in vitro*, notably the rate of transition e from infected non-productive bacteria to productively infected bacteria (eg, $\log(2)/e$ is the median eclipse phase duration), and the burst size b , *i.e.* the number of virions released by a lysed bacteria. The median eclipse phase time and the burst size were estimated from non-linear mixed effect models performed with the SAEM algorithm implemented in Monolix software (version 2018R2) (Kuhn and Lavielle, 2005) using data from one-step-growth experiments (see Supplementary material).

Presentation of the murine model of pulmonary phage therapy

The *E. coli* strain 536 is a virulent clinical strain belonging to the B2 phylogenetic group (Brzuszkiewicz et al., 2006). The *E. coli* strain 536-lux was obtained by the transformation of wild-type 536 strain with plasmid pCM17, which carries the *luxCDABE* operon from *Photobacterium luminescens* under *OmpC* promoter (Morin and Kaper, 2009). This plasmid is stable over 7 days without adding antibiotic in the medium (Dufour et al., 2015).

Phage 536_P1 (Dufour et al., 2015) is a virulent *Myoviridae* (149.4 Kb) that was amplified on strain 536 and purified according to molecular biology protocols including an endotoxin removal step (Endotrap Blue, Hyglos, Germany). The stock solution of $8.2 \log_{10}$ PFU/mL has an endotoxin concentration of 2.15 EU/mL.

Pneumonia was induced by intratracheal (IT) administration of strain 536-lux in eight-week-old BALB/cJRj anesthetized male mice (mixture of Ketamine and Xylazine administered intraperitoneally, IP). Exponentially growing bacteria in Luria Bertoni Broth (Lennox) at 37°C were centrifuged and washed twice and finally resuspended in PBS to obtain inoculums ranging from 7.6 to $8.2 \log_{10}$ CFU in 20 μ L.

Phage treatment was applied under general anesthesia (isoflurane, inspired fraction=2%) using IT or intravenous (IV, retro-orbital injection) administration two hours after infection. A volume of 20 μ L of phage 536_P1 was used, containing either a low (Lo; $6.6 \log_{10}$ PFU), medium (Me; $7.6 \log_{10}$ PFU) or high (Hi; $8.6 \log_{10}$ PFU) amount of phages.

At 2, 4, 8, 24 and 48h after phage administration, mice were sacrificed, and blood (in heparin coated tubes) and lungs were recovered and put on ice. Lungs were weighted, homogenized in PBS with protease inhibitor (EDTA free micro Tabs, SigmaAldrich). Lungs and blood were

then centrifuged (5000g, 15 min, 4°C). The supernatant was collected, serially diluted and spotted (in triplicate) on agar plates covered with strain 536 to titer phage 536_P1.

Bioluminescence and longitudinal follow up of bacterial load in lungs

In all our experiments photon emission of the luminescent bacteria was recorded at regular intervals under general anesthesia (isoflurane, inspired fraction=2%) using IVIS Spectrum and analyzed with Living Image software (PerkinElmer, Richmond, CA) as previously described (Debarbieux et al., 2010). Photons emitted within the chest area of each mouse were collected as photon/s/cm²/steradian.

To fully exploit the *in vivo* imaging tool we established the correlation between bioluminescence and CFU count in lungs using data from 17 untreated animals that were euthanized at different time points and for which lung bacteria was quantified by plating (Figure S1). The relationship obtained allowed us to derive from the luminescence signal an equivalent counts of CFU per gram of lung (noted CFUeq/g) that was used to obtain in all infected animals (untreated or treated by phages) the bacterial densities.

Phage distribution

A dose of 8 log₁₀ PFU of phage 536_P1 was administered in uninfected mice (N=53) using IT (N=16) or IV (N=37) route. At 2, 4, 8, 24 and 48h after phage administration, mice were sacrificed, and blood (in heparin coated tubes) and lungs were recovered and put on ice. Lungs were weighted, homogenized in PBS with protease inhibitor (EDTA free micro Tabs, SigmaAldrich). Lungs and blood were then centrifuged (5000g, 15 min, 4°C). The supernatant was collected, serially diluted and spotted (in triplicate) on agar plates covered with strain 536 to titer phage 536_P1.

We used the following standard pharmacokinetic model to estimate phage biodistribution:

$$\frac{dV_B}{dt} = -k_a V_B \quad (1)$$

$$\frac{dV}{dt} = F k_a V_B - k_e V \quad (2)$$

Where V and V_B denote phage quantity in lung and blood compartments, respectively, k_e is the elimination rate from lungs, k_a is the transfer rate from blood to lungs, F is the bioavailability in lungs after IV administration. The initial condition of these equations are $V_B(0) = dose$ and

$V(0) = 0$ in case of IV administration and $V_B(0) = 0$ and $V(0) = dose$ in case of IT administration. V is divided by the phage distribution weight (W) to obtain the phage concentration in lung compartment. The parameter W was deduced from the extrapolation of the median phage concentration at baseline ($8 \log_{10}$ PFU/g) following the IT administration of a single dose of $8 \log_{10}$ PFU, and F was obtained by performing the ratio of IV/IT area under the curves in the lung. Estimations of the other parameters were performed by naïve pooling of data (i.e. considering all the measures for each group (IV and IT) as if they came from a single individual and performing a nonlinear regression).

Bacterial kinetics in untreated mice

We characterize the natural bacterial kinetics (i.e., in absence of treatment) in animals infected with doses of $7.6 \log_{10}$ CFU (34 mice) or $8.2 \log_{10}$ CFU (12 mice) and the bacterial kinetics obtained by bioluminescence (see above) using the following model (one mouse was not analyzed due to death during bacterial inoculation, i.e. $N=45$):

$$\frac{dB}{dt} = \alpha \left(1 - \frac{B}{B_{max}} \right) B - \delta_M \frac{B}{B_{50} + B} \quad (3)$$

Where α is the exponential growth rate of bacteria (eg, $\log(2)/\alpha$ is the doubling time) and B_{max} is the maximal carrying capacity. The initial condition of B was estimated by the parameter B_0 ($B(0) = 10^{B_0}$). Consistent with previous work (Dufour et al., 2019), we assumed that in the first 3 hours of infection the immune system had no activity (eg, bacteria could grow exponentially during this period). After this period the elimination of bacteria by the host was modeled by a non-linear saturable term, given by $\delta_M \frac{B}{B_{50} + B}$. The model parameters were estimated in non-linear mixed effect models performed with the SAEM algorithm implemented in Monolix software (version 2018R2) (Kuhn and Lavielle, 2005).

Phage-bacteria interaction in treated mice

We characterized phage/bacterial interactions using data obtained in 142 mice infected by $7.6 \log_{10}$ CFU and treated subsequently (four mice were not analyzed due to death during bacterial inoculation, i.e. $N=138$), using the protocol described above, with a single dose (low, medium or high) of phages administered using IT or IV route (detail of the groups in Figure 1). Bacterial kinetics was derived from bioluminescence (see above) and lungs were collected at 2, 4, 8, 24 or 48h after phage administration to count PFU. We also considered data in untreated mice (see previous section) for the analysis.

The model describing interactions between phage 536_P1 and strain 536 relied on a system of ordinary differential equations (ODE) (Figure 1, Equations 4 to 10, presented below). We combined standard model of viral (Perelson and Guedj, 2015) and bacterial kinetics (Friberg and Guedj, 2020) to tease out the main parameters governing phage-bacteria interaction in mice. In lungs, phages can infect susceptible bacteria (S) following a mass-action term, with rate β . We also considered a population of refractory bacteria (of initial proportion f), noted R, that includes both phage-resistant and phage-inaccessible bacteria.

$$\frac{dS}{dt} = \alpha \left(1 - \frac{S + I_1 + I_2 + R}{B_{max}} \right) S - \delta_M \frac{S}{B_{50} + S + I_1 + I_2 + R} - \beta VS \quad (4)$$

$$\frac{dI_1}{dt} = \beta VS - eI_1 - \delta_M \frac{I_1}{B_{50} + S + I_1 + I_2 + R} \quad (5)$$

$$\frac{dI_2}{dt} = eI_1 - \delta_V I_2 - \delta_M \frac{I_2}{B_{50} + S + I_1 + I_2 + R} \quad (6)$$

$$\frac{dR}{dt} = \alpha \left(1 - \frac{S + I_1 + I_2 + R}{B_{max}} \right) R - \delta_M \frac{R}{B_{50} + S + I_1 + I_2 + R} \quad (7)$$

$$\frac{dV_B}{dt} = -k_a V_B \quad (8)$$

$$\frac{dV}{dt} = -\beta VS + b\delta_V I_2 + Fk_a V_B - k_e V \quad (9)$$

We considered the following initial condition for this system: $S(0) = 10^{B_0}(1 - f)$, $I_1(0) = 0$, $I_2(0) = 0$, $R(0) = 10^{B_0} \times f$, $V_B(0) = 0$, $V(0) = 0$. The dose of phages is added to V or V_B at the phage administration time by IT or IV route, respectively.

The model parameters in bold were estimated in non-linear mixed effect models performed with the SAEM algorithm implemented in Monolix software using data from all infected animals (version 2018R2) (Kuhn and Lavielle, 2005). All other parameters were fixed to the values determined above, either *in vitro* (e , b), in uninfected treated mice (k_a , F , k_e) or in infected untreated mice (α , B_{max} , δ_M , B_{50}).

Impact of bacterial kinetics on survival

Finally, we aimed to take into account the potential bias due to the fact that more severely infected animals died earlier, making the population study less and less representative as time goes by. In this approach, called joint models (Desmee et al., 2015), the probability to survive up to time t , $S(t)$, is modeled as: $S(t) = e^{-\int_0^t h(t)dt}$ where

$$h(t) = \frac{\kappa}{\lambda} \left(\frac{t}{\lambda}\right)^{\kappa-1} e^{\gamma \times \log(S(t)+I_1(t)+I_2(t)+R(t))} \quad (10)$$

$h(t)$ is the instantaneous hazard function, κ and λ are scale parameters characterizing the risk of death in absence of infection, and γ quantifies the impact of bacterial load on mortality. Thus, in this model $\gamma = 0$ is indicative of an absence of association between bacterial load and mortality, while $\gamma > 0$ indicates that increased bacterial load is associated with an increased risk of death.

Model evaluation

Individual fits were investigated to evaluate the different models. For the final model, we also evaluated the model predictions i) for the bacterial kinetics by using prediction corrected-visual predictive checks (Bergstrand et al., 2011), accounting for the risk of death and ii) visual predictive checks of the survival. Simulations for validation methods were made using mlxR package.

Model based simulations for the impact of treatment procedures and phage parameter values

Model based predictions were performed to investigate various scenarios as follows. For each simulation, the median profile of different model compartments (S, R, I_1, I_2 and V) was investigated using the parameters determined in Table 1 and different experimental conditions regarding the route of administration (IV or IT), the dose of phage administered (Lo: 5.6, Med: 7.6 or Hi: 8.6 \log_{10} PFU) and initial bacterial load (6, 7 or 8 \log_{10} CFUeq/g). We also evaluated the effects of different burst size (10 fold lower, 10 fold larger than the estimated value *in vitro*), using a single initial bacterial load of 7 \log_{10} CFUeq/g. Furthermore, when setting a burst size 10 times lower than the estimated value (and a fixed initial bacterial load of 7 \log_{10} CFUeq/g), we investigated the effect of three different lysis rates values: the value estimated *in vivo* (0.09 h^{-1}); a value three times higher (0.27 h^{-1}); a value estimated *in vitro* (1.39 h^{-1}).

Ethical statement

A total of 241 eight-week-old BALB/cJRj male mice (Janvier, France) were housed in animal facility in accordance with French and European regulations on the care and protection of laboratory animals. Food and drink were provided ad libitum. Protocol was approved by the veterinary staff of the Institut Pasteur animal facility (approval number 20.173) and the National

Ethics Committee regulating animal experimentation (APAFIS#26874-2020081309052574 v1).

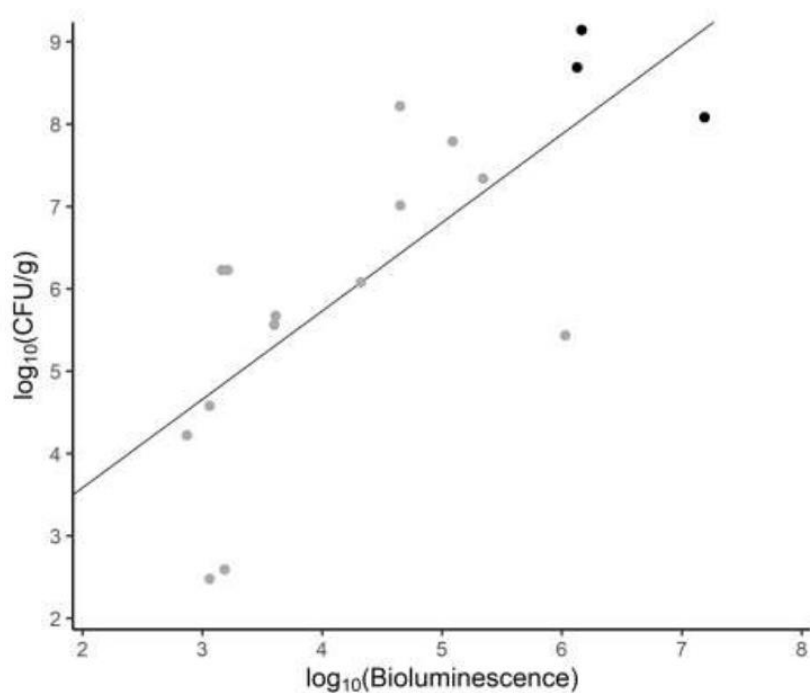
Authors contribution

JDR, JG, LD and RD, conceived the study and secured funding. BG, FH, RD and RK performed experiments. JDR, JG, JS, LD and RD analyzed the data. JG, JS and TTN performed modeling. ND provided feedback. JS and RD wrote the first draft. JG and LD reviewed and edited final draft.

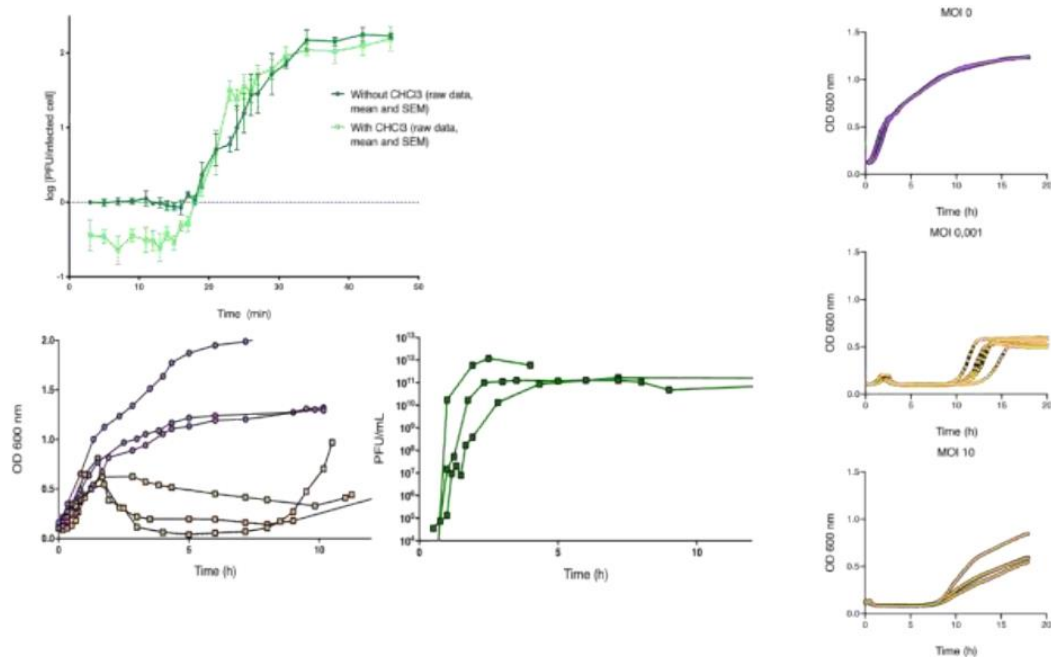
Declaration of interests

Authors declare no competing interest.

Supplementary figures

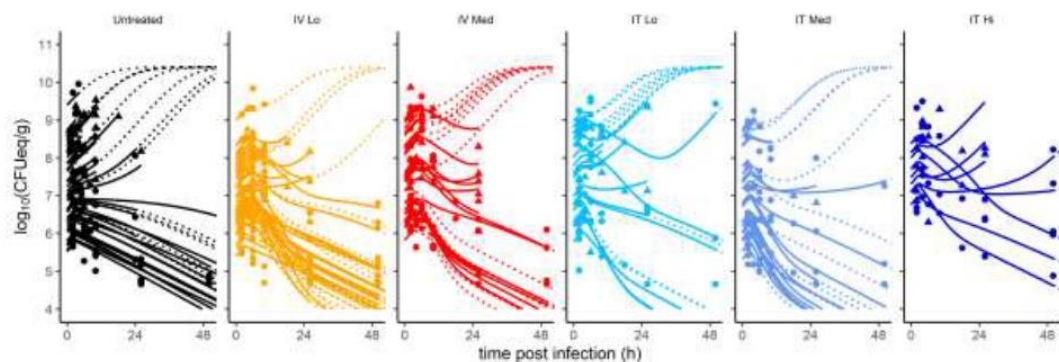


Supplementary figure S1. Correlation between bioluminescence and CFU from lungs of infected and untreated mice. The regression line equation is $\log_{10}(CFU/g) = 1.44 + 1.07 \log_{10}(Bioluminescence)$. From 17 mice, 14 were infected with an inoculum of 7.6 log₁₀ CFU (grey dots) and 3 with an inoculum of 8.2 log₁₀ CFU (black dots).

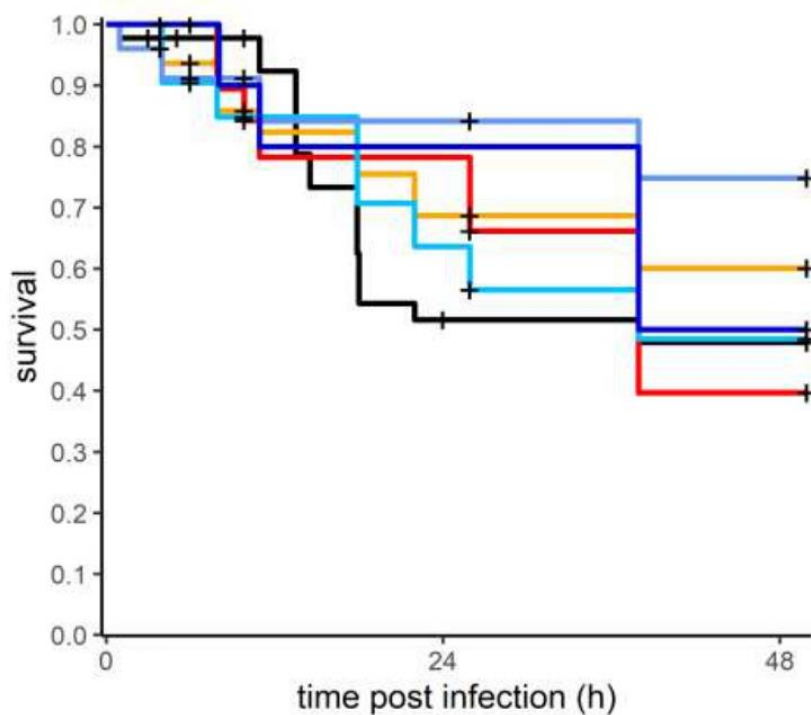


Supplementary figure S2. *In vitro* experiments describing the dynamic interactions between strain 536 and phage 536_P1.

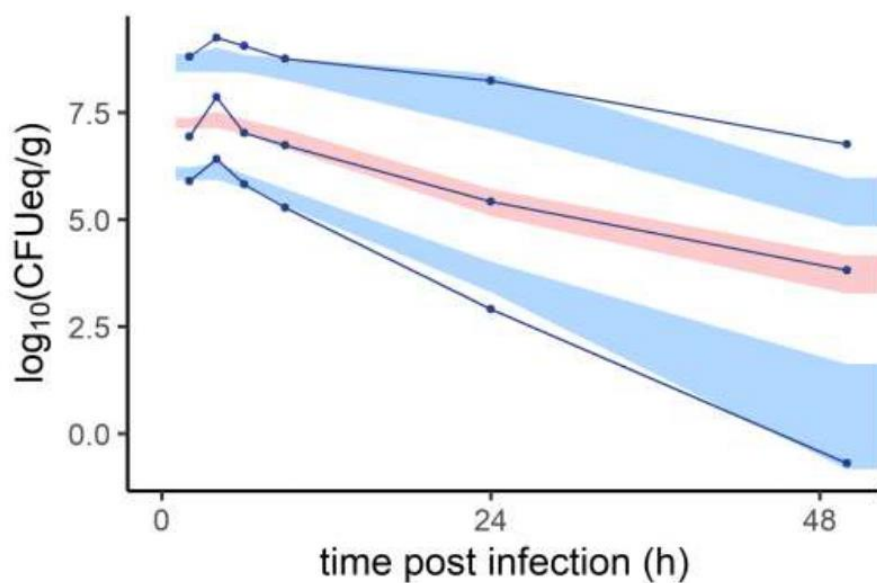
The one step growth data (3 replicates) are presented on top left, the dynamic quantification of strain 536 and phage 536_P1 (3 replicates at MOI 0 and at MOI 0.001) are presented on bottom left, and the kinetics of lysis (3 or 4 replicates at MOI 0, MOI 0.0001 and MOI 10) are presented on the right (see Supplementary methods).



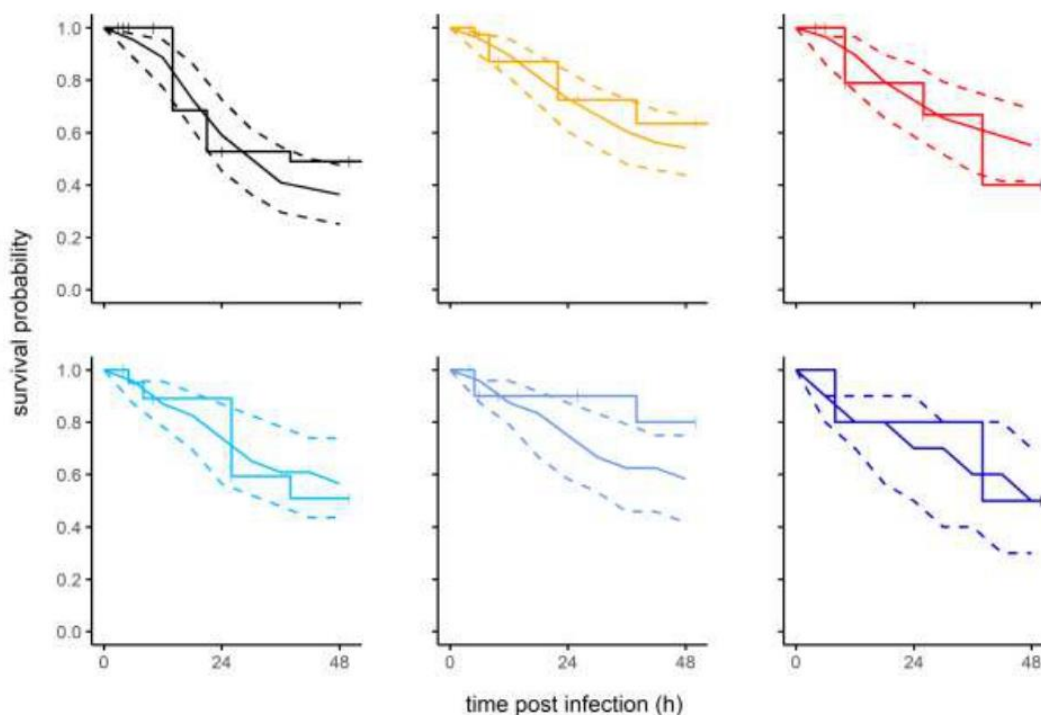
Supplementary figure S3. Longitudinal bacterial load data and individual fits from all infected mice (N=183). Solid lines represent individual predictions from the infection time to the last observation time. The dotted lines are individual predictions starting from the sacrifice time point. Interrupted solid lines correspond to animal that die before the scheduled sacrifice time. Circles correspond to live animals while triangles correspond to animals at time of sacrifice after recording the bioluminescence.



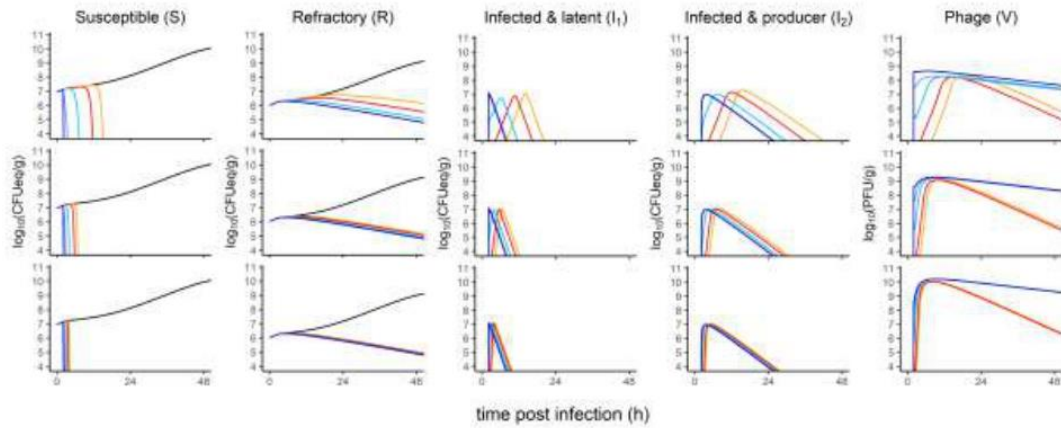
Supplementary figure S4. Survival curves over time of *E. coli* strain 536 infected mice and untreated (black, N=45), or treated by phage 536_P1 administered intravenously (IV) at low dose of $5.6 \log_{10}$ PFU (orange, N=50), medium dose of $7.6 \log_{10}$ PFU (red, N=29), intratracheally (IT) at low dose (cyan, N=24), medium dose (blue, N=25) or high dose of $8.6 \log_{10}$ PFU (dark blue, N=10). Crosses represent sacrifice time points.



Supplementary figure S5. Prediction-corrected visual predictive check plot for the bacterial load, using the final model. The 90% confidence intervals around the 10th (bottom), 50th (middle) and 90th (top) model-predicted percentiles, obtained from 1000 simulated data sets from the model, are represented as blue (10th and 90th percentiles) and pink (50th percentile) areas. The blue dots represent the 10th, 50th and 90th percentiles of the observations from the original data set.



Supplementary figure S6. Visual predictive check plot for the survival in each group, using the final model. Untreated mice are in black, phage 536_P1 treated mice (2 h after infection) groups are in colors, with different phage administration scheme: intravenously (IV) at low dose ($5.6 \log_{10}$ PFU) in orange or medium dose ($7.6 \log_{10}$ PFU) in red, intratracheally (IT) at low dose in cyan, medium dose in blue or high dose ($8.6 \log_{10}$ PFU) in dark blue (IT Hi). The dashed lines around the solid line represent the 10th (bottom), 50th (middle) and 90th (top) percentiles of survival obtained from 1000 simulated data sets from the model. The stepped line is the observed survival (Kaplan-Meier curve), with crosses representing censoring.



Supplementary figure S7. Median profile of model predictions for different burst sizes: 57, 570 (*i.e.* the estimated value) and 5700 PFU/CFU in the top, middle, and bottom line, respectively), with a fixed inoculum of $7 \log_{10}$ CFUeq/g. Each column represents different model compartments: bacteria (S, R, I_1 and I_2) and phages in lungs. The colors are indicative of both the route of administration (IT or IV) and the dose of phage received (Lo: $5.6 \log_{10}$ PFU; Med: $7.6 \log_{10}$ PFU; Hi: $7.6 \log_{10}$ PFU): untreated (black), IV-Lo (orange), IV-Med (red), IT-Lo (cyan), IT-med (blue), IT-high (dark blue).

Supplementary Material

Purification of bacteriophages

Phage 536_P1 was amplified on strain 536 and purified by first, filtration of lysate to 0.1µm followed by tangential ultrafiltration (Vivaflow™, Sartorius, 100kDa), two ultracentrifugations on cesium chloride, dialysed against TN (Tris-HCl pH8, NaCl 150 mM), passed three times through an endotoxin removal column (Endotrap Blue, Hyglos, Germany) and filtered-sterilized at 0.22µm.

***In vivo* experiments**

Intratracheal administration (20µL) was performed in anesthetized mice using elongated tips guided by the glottis visualization with a cold light. The correct position of the tip was checked by the liquid movement inside the tip caused by the ventilatory movement of the animal. The liquid was then expelled from the tip with a pipette, and 100 µL of air was then pushed to aerosolize the solution to the bronchoalveolar space. Mice were maintained in prone position during at least 5 minutes.

***In vitro* experiments**

One-Step-Growth experiments

The one-step growth experiment was performed in triplicate, using Lysogeny Broth (LB, Difco Bacto-Tryptone: 10 g/L, Difco Yeast extract Difco 5 g/L, NaCl 5 g/L). As previously described (Hyman and Abedon, 2009), an exponential-phase culture (9 ml) of strain 536 (optical density at 600 nm 0,25, *ie* 9 log₁₀ CFU) was mixed with 1mL of 8 log₁₀ PFU of phage 536_P1 (MOI of 0.1). The mixture was placed under constant shaking (100 rpm) at 37°C. Duplicated samples were collected at 2-min intervals during 30 minutes. Six minutes after mixing phages and bacteria, the mixture was 1000 fold diluted in LB medium in order to lower bacterial concentration and avoid any reinfection. Out of the duplicated samples, one was immediately diluted and plated for phage titration, while the second was treated with 10% (vol/vol) chloroform to release intracellular phages in order to determine the eclipse period before phage titration.

Dynamic quantification of bacteria and phages over time

We first established the correlation between optical density and CFU for strain 536 by plating on LB agar dilutions of independent aliquots of exponentially growing cells from which OD600 nm was recorded before dilutions. Then, an exponential bacterial culture (50 mL of 1.10⁷

CFU/mL (ie a total of $8.7 \log_{10}$ CFU) was mixed with $5.7 \log_{10}$ PFU (ie MOI 1.10^{-3}) and incubated under constant shaking (150 rpm, 37°C). Samples from this mixture were withdrawn at specific time points to measure optical density at 600 nm and count the number of phages after serial dilutions and plating on LB agar covered with strain 536. Experiments were performed three times from independent cultures.

Kinetic of lysis

To assess the behavior of strain 536 in presence of phage 536_P1 at different MOI, OD_{600 nm} was recorded over time from microtiter plates. In each well, 180 µL of an exponential culture of strain 536 (OD 0.1) was mixed with 20 µL of phage 536_P1 at different concentrations (each replicated six times) to reach MOI of 0.001; 0.01; 0.1; 1 and 10. Controls included wells with phage alone, or bacteria alone or LB alone. The microplates were incubated in a microplate reader (Tecan) set at 37°C during 12h with orbital agitation. Three independent experiments were performed.

***In vitro* data analysis**

Models of *in vitro* data were built using similar methodology used for analyzing *in vivo* PK/PD data: parameters of mixed models were estimated using SAEM algorithm and the model selection was based on BIC. Model of phage/bacteria interactions *in vitro* was built in three steps.

First, one step growth experiments reproduced a single lysis cycle of the phage and therefore allowed to estimate the burst size (*i.e.* the number of virions released from each bacterium lysed) and the time of a cycle, including an eclipse period. The best fit of these one step growth data was obtained using a multiple transit compartments model. The burst size (570 virions) was estimated with acceptable precision of estimation (relative standard error <30%). The median period for a phage to produce and release new phages was estimated at 35 minutes. Second, bacterial growth was characterized from the above *in vitro* experiments exploiting the condition MOI 0 (without any phage). The bacterial doubling time was estimated at 63 minutes from the model. Third, lysis kinetics experiments (at MOI > 0) were modelled using a predator-prey model with similar structure to *in vivo* model (see equations and table of parameters below). Burst size and bacterial growth were fixed at values respectively obtained at the above first and second steps. The lysis rate (corresponding to rate between infected latent and infected productive bacteria) was fixed from one step growth experiments on 536_P1. The eclipse period

was obtained by sensitivity analysis on the transition rate e . The proportion of phage-resistant clones at baseline and their relative fitness were characterized from this model.

$$\begin{aligned}\frac{dS}{dt} &= \alpha \left(1 - \frac{S + I_1 + I_2 + R}{10^{B_{max}}}\right) S - \beta VS \\ \frac{dI_1}{dt} &= \beta VS - eI_1 \\ \frac{dI_2}{dt} &= eI_1 - \delta_V I_2 \\ \frac{dR}{dt} &= w\alpha \left(1 - \frac{S + I_1 + I_2 + R}{10^{B_{max}}}\right) R \\ \frac{dV}{dt} &= -\beta VS + \delta_V bI_2\end{aligned}$$

Table S1. Parameter estimates of the final *in vitro* model

Parameter	Name	Unit	Fixed/Estimated from	Fixed effect (rse, %)	Sd of random effect (rse, %)
Bacterial growth rate	α	min^{-1}	Coev + Lysis (MOI 0)	0.011 (-)	0 (-)
Initial number of susceptible	S_0	\log_{10} CFU mL ⁻¹	Coev + Lysis (MOI 0)	7.9 (-)	0 (-)
Bacterial load plateau	B_{max}	\log_{10} CFU mL ⁻¹	Coev + Lysis (MOI 0)	9.9 (-)	0 (-)
Infectivity rate	β	\log_{10} mL CFU ⁻¹ min ⁻¹	Lysis (MOI > 0)	9.6 (<1)	0.01 (26)
Transition rate	e	min^{-1}	Lysis (MOI > 0)	0.02 (-)	0 (-)
Lysis rate	δ_V	min^{-1}	Dufour et al., 2017	0.5 (-)	0 (-)
Burst size	b	PFU CFU ⁻¹	OSG	570 (-)	0 (-)
Initial number of resistant	R_0	\log_{10} CFU mL ⁻¹	Lysis (MOI > 0)	6.7 (<1)	0.05 (10)
Relative fitness	w	-	Lysis (MOI > 0)	0.49 (4)	0.27 (10)
Additive error on bacterial load	σ_B	\log_{10} PFU g ⁻¹	Lysis (MOI > 0)	-	0.1 (1)

Sd: standard deviation, rse: relative standard error

References

- Alemayehu, D., Casey, P.G., McAuliffe, O., Guinane, C.M., Martin, J.G., Shanahan, F., Coffey, A., Ross, R.P., and Hill, C. (2012). Bacteriophages phiMR299-2 and phiNH-4 can eliminate *Pseudomonas aeruginosa* in the murine lung and on cystic fibrosis lung airway cells. *mBio* **3**, e00029-00012.
- Aslam, S., Lampley, E., Wooten, D., Karris, M., Benson, C., Strathdee, S., and Schooley, R.T. (2020). Lessons Learned From the First 10 Consecutive Cases of Intravenous Bacteriophage Therapy to Treat Multidrug-Resistant Bacterial Infections at a Single Center in the United States. *Open forum infectious diseases* **7**, ofaa389.
- Bergeron, Y., Ouellet, N., Deslauriers, A.M., Simard, M., Olivier, M., and Bergeron, M.G. (1998). Cytokine kinetics and other host factors in response to pneumococcal pulmonary infection in mice. *Infect Immun* **66**, 912-922.
- Bergstrand, M., Hooker, A.C., Wallin, J.E., and Karlsson, M.O. (2011). Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *AAPS J* **13**, 143-151.
- Brzuszkiewicz, E., Bruggemann, H., Liesegang, H., Emmerth, M., Otschlager, T., Nagy, G., Albermann, K., Wagner, C., Buchrieser, C., Emody, L., *et al.* (2006). How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc Natl Acad Sci U S A* **103**, 12879-12884.
- Cairns, B.J., Timms, A.R., Jansen, V.A., Connerton, I.F., and Payne, R.J. (2009). Quantitative models of in vitro bacteriophage-host dynamics and their application to phage therapy. *PLoS Pathog* **5**, e1000253.
- Carmody, L.A., Gill, J.J., Summer, E.J., Sajjan, U.S., Gonzalez, C.F., Young, R.F., and LiPuma, J.J. (2010). Efficacy of bacteriophage therapy in a model of *Burkholderia cenocepacia* pulmonary infection. *J Infect Dis* **201**, 264-271.
- Corbellino, M., Kieffer, N., Kutateladze, M., Balarjishvili, N., Leshkasheli, L., Askilashvili, L., Tsertsvadze, G., Rimoldi, S.G., Nizharadze, D., Hoyle, N., *et al.* (2020). Eradication of a Multidrug-Resistant, Carbapenemase-Producing *Klebsiella pneumoniae* Isolate Following Oral and Intra-rectal Therapy With a Custom Made, Lytic Bacteriophage Preparation. *Clin Infect Dis* **70**, 1998-2001.
- Dabrowska, K., and Abedon, S.T. (2019). Pharmacologically Aware Phage Therapy: Pharmacodynamic and Pharmacokinetic Obstacles to Phage Antibacterial Action in Animal and Human Bodies. *Microbiol Mol Biol Rev* **83**.
- Debarbieux, L., Leduc, D., Maura, D., Morello, E., Criscuolo, A., Grossi, O., Balloy, V., and Touqui, L. (2010). Bacteriophages can treat and prevent *Pseudomonas aeruginosa* lung infections. *J Infect Dis* **201**, 1096-1104.
- Desmee, S., Mentre, F., Veyrat-Follet, C., and Guedj, J. (2015). Nonlinear Mixed-effect Models for Prostate-specific Antigen Kinetics and Link with Survival in the Context of Metastatic Prostate Cancer: A Comparison by Simulation of Two-stage and Joint Approaches. *AAPS J* **17**, 691-699.
- Dion, M.B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nat Rev Microbiol* **18**, 125-138.
- Dufour, N., Debarbieux, L., Fromentin, M., and Ricard, J.D. (2015). Treatment of Highly Virulent Extraintestinal Pathogenic *Escherichia coli* Pneumonia With Bacteriophages. *Crit Care Med* **43**, e190-198.
- Dufour, N., Delattre, R., Chevallereau, A., Ricard, J.D., and Debarbieux, L. (2019). Phage Therapy of Pneumonia Is Not Associated with an Overstimulation of the Inflammatory Response Compared to Antibiotic Treatment in Mice. *Antimicrob Agents Chemother* **63**.

- Friberg, L.E., and Guedj, J. (2020). Acute bacterial or viral infection-What's the difference? A perspective from PKPD modellers. *Clin Microbiol Infect* 26, 1133-1136.
- Gordillo Altamirano, F.L., and Barr, J.J. (2019). Phage Therapy in the Postantibiotic Era. *Clin Microbiol Rev* 32.
- Henry, M., Lavigne, R., and Debarbieux, L. (2013). Predicting in vivo efficacy of therapeutic bacteriophages used to treat pulmonary infections. *Antimicrob Agents Chemother* 57, 5961-5968.
- Hyman, P., and Abedon, S.T. (2009). Practical methods for determining phage growth parameters. *Methods Mol Biol* 501, 175-202.
- Jault, P., Leclerc, T., Jennes, S., Pimay, J.P., Que, Y.A., Resch, G., Rousseau, A.F., Ravat, F., Carsin, H., Le Floch, R., *et al.* (2019). Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by *Pseudomonas aeruginosa* (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial. *Lancet Infect Dis* 19, 35-45.
- Jennes, S., Merabishvili, M., Soentjens, P., Pang, K.W., Rose, T., Keersebilck, E., Soete, O., Francois, P.M., Teodorescu, S., Verween, G., *et al.* (2017). Use of bacteriophages in the treatment of colistin-only-sensitive *Pseudomonas aeruginosa* septicaemia in a patient with acute kidney injury-a case report. *Crit Care* 21, 129.
- Jiang, W., Yan, Y., Ji, W., Wang, Y., and Chen, Z. (2014). Clinical significance of different bacterial load of *Mycoplasma pneumoniae* in patients with *Mycoplasma pneumoniae* pneumonia. *Braz J Infect Dis* 18, 124-128.
- Kasman, L.M., Kasman, A., Westwater, C., Dolan, J., Schmidt, M.G., and Norris, J.S. (2002). Overcoming the phage replication threshold: a mathematical model with implications for phage therapy. *J Virol* 76, 5557-5564.
- Kortright, K.E., Chan, B.K., Koff, J.L., and Turner, P.E. (2019). Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. *Cell Host Microbe* 25, 219-232.
- Kuhn, E., and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis* 49, 1020-1038.
- Leitner, L., Ujmajuridze, A., Chanishvili, N., Goderdzishvili, M., Chkonia, I., Rigvava, S., Chkhotua, A., Changashvili, G., McCallin, S., Schneider, M.P., *et al.* (2020). Intravesical bacteriophages for treating urinary tract infections in patients undergoing transurethral resection of the prostate: a randomised, placebo-controlled, double-blind clinical trial. *Lancet Infect Dis*.
- Leung, C.Y.J., and Weitz, J.S. (2017). Modeling the synergistic elimination of bacteria by phage and the innate immune system. *J Theor Biol* 429, 241-252.
- Lin, Y., Quan, D., Yoon Kyung Chang, R., Chow, M.Y.T., Wang, Y., Li, M., Morales, S., Britton, W.J., Kutter, E., Li, J., *et al.* (2020). Synergistic activity of phage PEV20-ciprofloxacin combination powder formulation-A proof-of-principle study in a *P. aeruginosa* lung infection model. *European journal of pharmaceuticals and biopharmaceutics : official journal of Arbeitsgemeinschaft für Pharmazeutische Verfahrenstechnik eV*.
- Lourenco, M., Chaffringeon, L., Lamy-Besnier, Q., Pedron, T., Campagne, P., Eberl, C., Berard, M., Stecher, B., Debarbieux, L., and De Sordi, L. (2020). The Spatial Heterogeneity of the Gut Limits Predation and Fosters Coexistence of Bacteria and Bacteriophages. *Cell Host Microbe* 28, 390-401 e395.
- Luong, T., Salabarria, A.C., and Roach, D.R. (2020). Phage Therapy in the Resistance Era: Where Do We Stand and Where Are We Going? *Clin Ther* 42, 1659-1680.
- Marriott, H.M., and Dockrell, D.H. (2007). The role of the macrophage in lung disease mediated by bacteria. *Exp Lung Res* 33, 493-505.
- Marshall, S., Madabushi, R., Manolis, E., Krudys, K., Staab, A., Dykstra, K., and Visser, S.A.G. (2019). Model-Informed Drug Discovery and Development: Current Industry Good Practice

- and Regulatory Expectations and Future Perspectives. *CPT: pharmacometrics & systems pharmacology* 8, 87-96.
- Melo, L.D.R., Oliveira, H., Pires, D.P., Dabrowska, K., and Azeredo, J. (2020). Phage therapy efficacy: a review of the last 10 years of preclinical studies. *Crit Rev Microbiol* 46, 78-99.
- Morello, E., Sausseureau, E., Maura, D., Huerre, M., Touqui, L., and Debarbieux, L. (2011). Pulmonary bacteriophage therapy on *Pseudomonas aeruginosa* cystic fibrosis strains: first steps towards treatment and prevention. *PLoS One* 6, e16963.
- Morin, C.E., and Kaper, J.B. (2009). Use of stabilized luciferase-expressing plasmids to examine in vivo-induced promoters in the *Vibrio cholerae* vaccine strain CVD 103-HgR. *FEMS Immunol Med Microbiol* 57, 69-79.
- Oechslin, F., Piccardi, P., Mancini, S., Gabard, J., Moreillon, P., Entenza, J.M., Resch, G., and Que, Y.A. (2017). Synergistic Interaction Between Phage Therapy and Antibiotics Clears *Pseudomonas Aeruginosa* Infection in Endocarditis and Reduces Virulence. *J Infect Dis* 215, 703-712.
- Payne, R.J., and Jansen, V.A. (2001). Understanding bacteriophage therapy as a density-dependent kinetic process. *J Theor Biol* 208, 37-48.
- Payne, R.J., and Jansen, V.A. (2003). Pharmacokinetic principles of bacteriophage therapy. *Clin Pharmacokinet* 42, 315-325.
- Perelson, A.S., and Guedj, J. (2015). Modelling hepatitis C therapy--predicting effects of treatment. *Nat Rev Gastroenterol Hepatol* 12, 437-445.
- Pouillot, F., Chomton, M., Blois, H., Courroux, C., Noelig, J., Bidet, P., Bingen, E., and Bonacorsi, S. (2012). Efficacy of bacteriophage therapy in experimental sepsis and meningitis caused by a clone O25b:H4-ST131 *Escherichia coli* strain producing CTX-M-15. *Antimicrob Agents Chemother* 56, 3568-3575.
- Roach, D.R., Leung, C.Y., Henry, M., Morello, E., Singh, D., Di Santo, J.P., Weitz, J.S., and Debarbieux, L. (2017). Synergy between the Host Immune System and Bacteriophage Is Essential for Successful Phage Therapy against an Acute Respiratory Pathogen. *Cell Host Microbe* 22, 38-47 e34.
- Sarker, S.A., Sultana, S., Reuteler, G., Moine, D., Descombes, P., Charton, F., Bourdin, G., McCallin, S., Ngom-Bru, C., Neville, T., *et al.* (2016). Oral Phage Therapy of Acute Bacterial Diarrhea With Two Coliphage Preparations: A Randomized Trial in Children From Bangladesh. *EBioMedicine* 4, 124-137.
- Schooley, R.T., Biswas, B., Gill, J.J., Hernandez-Morales, A., Lancaster, J., Lessor, L., Barr, J.J., Reed, S.L., Rohwer, F., Benler, S., *et al.* (2017). Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant *Acinetobacter baumannii* Infection. *Antimicrob Agents Chemother* 61.
- Smith, A.M., McCullers, J.A., and Adler, F.R. (2011). Mathematical model of a three-stage innate immune response to a pneumococcal lung infection. *J Theor Biol* 276, 106-116.
- Smith, H.W., Huggins, M.B., and Shaw, K.M. (1987). The control of experimental *Escherichia coli* diarrhoea in calves by means of bacteriophages. *J Gen Microbiol* 133, 1111-1126.
- Sousa, J.A.M., and Rocha, E.P.C. (2019). Environmental structure drives resistance to phages and antibiotics during phage therapy and to invading lysogens during colonisation. *Sci Rep* 9, 3149.
- Weld, R.J., Butts, C., and Heinemann, J.A. (2004). Models of phage growth and their applicability to phage therapy. *J Theor Biol* 227, 1-11.

Quatrième partie : Discussion

La première partie des travaux de recherche de ce doctorat est consacrée à l'étape se trouvant en amont de la modélisation par MNLEM : la planification du protocole. Dans la seconde partie de cette thèse, les MNLEM sont utilisés pour analyser des données de phagothérapie. Ces deux parties sont discutées successivement.

4.1. Protocoles en modélisation

Un protocole d'essai clinique bien choisi permet d'obtenir une bonne précision d'estimation des paramètres du modèle et donc de pouvoir répondre aux objectifs de l'étude. Deux projets de ce doctorat s'inscrivent dans le cadre du développement de méthodes pour évaluer et optimiser des protocoles à partir de la FIM. Un autre travail, plus appliqué, porte sur l'évaluation de protocoles par simulations en oncologie.

4.1.1. Protocoles robustes prenant en compte l'incertitude du modèle

Dans le premier travail, nous avons développé et évalué une approche de protocoles robustes tenant compte de l'incertitude du modèle, dans le cadre d'études longitudinales avec réponse binaire. Cette stratégie de planification de protocoles est basée sur la FIM, évaluée par MC/HMC. Cette approche permet de prendre en compte l'incertitude du choix de modèle en supposant plusieurs modèles de structure candidats, à partir d'un critère d'optimalité composé (Atkinson et al. 2007). Par la même occasion, nous avons proposé la possibilité de considérer un sous-ensemble de paramètres dont on cherche à maximiser la précision d'estimation, sans trop pénaliser celle des autres paramètres. Si ce sous-ensemble de paramètres contient des effets de covariables, optimiser un protocole par cette méthode aura notamment pour but d'optimiser la puissance du test de Wald pour ces effets de covariables. Cette approche de planification de protocole robuste a été évaluée par simulations pour la première fois. Cette analyse a confirmé la pertinence de cette méthode, le protocole robuste s'étant montré plus performant que les autres protocoles évalués (un protocole non-optimisé et un protocole optimisé par une approche non robuste).

Concernant la méthode permettant de s'intéresser davantage à certains paramètres au travers du critère DDs, nous avons proposé une méthode permettant d'obtenir le meilleur compromis

possible entre les précisions d'estimation des deux sous-ensembles de paramètres (ψ_S et ψ_T , respectivement les paramètres d'intérêt et les autres paramètres estimés). Notre proposition est de maximiser le produit des efficacités D et D_S , cette méthode est cependant adaptable selon l'objectif de l'étude. Nous nous sommes placés dans un cas où ψ_S était identique pour chacun des modèles candidats, il serait intéressant de réaliser un autre travail avec des ψ_S de différentes tailles selon le modèle candidat considéré, par exemple des modèles PKPD avec des relations concentration-effet différentes (linéaires et non-linéaires).

Le biais d'estimation obtenu lors de notre évaluation par simulations est un élément de discussion important. Différents algorithmes d'estimation, réglages d'algorithmes et différents logiciels ont été essayés afin de parvenir aux résultats les moins biaisés possibles pour l'ensemble des paramètres des différents modèles, mais un léger biais d'estimation demeure, notamment pour l'effet traitement et les paramètres du modèle exponentiel (modèle le plus complexe parmi les modèles considérés). Il serait intéressant d'effectuer une comparaison d'algorithmes d'estimation pour les modèles logistiques mixtes.

Dans l'application choisie dans ce travail, nous avons choisi de simuler des données avec un modèle parmi les autres modèles candidats. Il serait intéressant de simuler des données avec un modèle plus complexe, non considéré *a priori* lors de la planification du protocole (puis utilisé ou non lors de l'analyse des données), afin d'étudier l'apport de l'approche robuste dans un contexte plus proche de la réalité. Dans le but d'étudier la sensibilité de notre méthode, d'autres scénarios pourraient être considérés, avec plus ou moins de modèles candidats, des poids différents sur chaque modèle ou encore des problématiques de protocole plus complexes, par exemple avec davantage de temps de mesures à optimiser que dans l'application présentée. L'approche de protocole robuste est proposée dans le cadre de données binaires répétées. Cette stratégie pourrait néanmoins être appliquée à d'autres types de données, continues comme non-continues. Par exemple, le même type d'approche, prenant en compte l'incertitude du modèle en utilisant la théorie du design composé, a été appliquée aux données de comptage analysées par modèles de Poisson, en tenant compte de l'incertitude des paramètres en plus de celle du modèle (Loingeville et al. 2019). Il reste cependant à évaluer la pertinence de cette approche étendant les protocoles robustes à l'incertitude des paramètres par CTS, voire à appliquer cette méthode à d'autres types de données.

Le principal pré-requis pour utiliser une approche de planification de protocole robuste est de disposer d'une méthode adéquate pour évaluer la FIM des modèles candidats. La méthode FO permet d'évaluer efficacement la FIM pour modèles de données continues (Mentré et al. 1997),

tandis que les méthodes MC/HMC et MC/AGQ, voire QRMC/AGQ sont préférables pour les données binaires et de comptage (Riviere et al. 2016; Ueckert and Mentré 2017). La FIM a été évaluée par QRMC/AGQ pour le travail sur l’algorithme multiplicatif (sections 2.2.1 et 4.1.2.), pour un exemple d’application similaire à celui utilisé pour le développement de l’approche robuste. De plus, il serait opportun de développer une méthode permettant d’évaluer la FIM pour les modèles conjoints, dont l’utilisation en pharmacométrie est de plus en plus fréquente (Desmée et al. 2015; Tardivon et al. 2019; Yu et al. 2020).

L’optimisation de protocole adaptatif par la FIM est aussi une approche qui mérite d’être développée. L’apport de cette approche a d’ailleurs été étudié pour une application en oncologie (Lestini et al. 2015). Ce travail avait montré qu’utiliser la FIM empirique, obtenue à partir de données issues d’une première étape, permettait de choisir efficacement le protocole de la deuxième étape. Cela permettait finalement d’aboutir à une estimation de paramètres plus précise et moins biaisée qu’avec un protocole à une étape, notamment lorsque les paramètres étaient mal choisis pour la planification de la première étape. Ce type d’approche est compatible avec l’approche robuste présentée dans cette thèse. Pour ce faire, la première étape est planifiée selon la méthode présentée dans l’article, avec plusieurs modèles candidats et un poids identique pour chacun des modèles définis, en tenant éventuellement compte de l’incertitude des paramètres de chaque modèle. A la fin de la première étape du protocole, les données obtenues par les M ($m = 1, \dots, M$) sont décrites par les modèles candidats. Afin de planifier la seconde étape, les poids de chacun des modèles w_m sont redéfinis à partir des AIC obtenues, selon le principe du *model averaging* (équation 36) :

$$w_m = \frac{e^{-\frac{AIC_m - AIC_{min}}{2}}}{\sum_{m=1}^M e^{-\frac{AIC_m - AIC_{min}}{2}}}, \quad (36)$$

avec AIC_m , l’AIC du modèle candidat m et AIC_{min} l’AIC du modèle le plus parcimonieux (avec l’AIC le plus bas). Les valeurs de paramètres sont aussi mises à jour pour la planification de la seconde étape de l’étude.

4.1.2. Algorithmes multiplicatifs

Dans l’article discuté ci-dessus (section 4.1.1.), la planification de protocole est réalisée selon une approche combinatoire (en évaluant tous les protocoles possibles puis en choisissant le meilleur). L’approche robuste est néanmoins compatible avec l’utilisation d’un algorithme

d'optimisation de protocole, dès lors que cet algorithme puisse converger vers le protocole optimal selon le critère d'optimalité utilisé.

Le second travail sur les protocoles consistait à étendre l'utilisation d'algorithmes multiplicatifs à l'optimisation de protocoles d'études longitudinales fondée sur la FIM pour les modèles mixtes. Cet algorithme possède de multiples avantages tels que sa simplicité de fonctionnement et d'intégration aux différents problèmes d'optimisation de protocole. Lors de l'utilisation d'algorithmes multiplicatifs, l'espace de design est exploré de manière discrète, à partir de contraintes de design énoncées avant calcul. A partir de ces contraintes sont donc énoncés les protocoles élémentaires possibles de manière combinatoire (de même que pour le premier travail présenté section 2.1.), avec une FIM associée à chaque protocole. Ce sont en fait les poids associés à chaque design élémentaire qui sont optimisés par l'algorithme. Les poids finaux sont alors les proportions de sujets à allouer aux designs élémentaires correspondants. Cet algorithme est applicable à différents types de données et de modèles, dès lors qu'une méthode permettant d'évaluer la FIM de manière adéquate est disponible. Nous avons illustré l'intérêt de cet algorithme dans deux exemples. Le premier était une optimisation de temps de mesure et de doses pour un modèle de réponse continue avec une FIM évaluée par FO. Le second était une optimisation de temps de mesure pour un modèle de données binaires répétées (similaire à l'un de ceux utilisés pour le premier travail au cours duquel la FIM était évaluée par MC/HMC), avec une FIM évaluée par QRMC/AGQ, une méthode récente permettant d'évaluer la FIM efficacement pour des modèles de réponses continues et discrètes et plus rapidement que par MC/HMC pour des modèles avec un nombre réduit d'effets aléatoires, ce qui était le cas ici (Ueckert and Mentré 2015).

Dans la première application, nous avons montré que l'utilisation de cet algorithme permet d'optimiser à la fois des temps de mesures et des doses. Cela pourrait être étendu à d'autres éléments de protocoles tels que le nombre de temps de mesure ou des modalités d'administration différentes.

Deux versions de l'algorithme ont été programmées, la première en langage R et la seconde en C. L'algorithme écrit en C a montré de meilleures performances en temps de calcul que celui en R pour les différentes applications étudiées. Par exemple, pour l'optimisation des poids des 1932 protocoles élémentaires différents dans la première application, le temps de calcul était de 31 s en R contre 432 s en C. L'algorithme écrit en C plutôt qu'en R devrait donc montrer un intérêt particulier lorsque le nombre protocoles élémentaires possibles est très grand, par exemple au-delà de 100 000.

Les preuves de monotonie et de convergence vers le protocole D-optimal ont été fournies précédemment, pour tout λ (le paramètre de calibration) compris entre 0 et 1 (Yu 2010). Proposer ces démonstrations détaillées dans le manuscrit pourrait faciliter sa lecture. Nous avons choisi de fixer le paramètre de calibration à 0.99, cette valeur nous ayant permis d'obtenir la convergence de l'algorithme pour nos différentes applications. La sensibilité de l'algorithme à ce paramètre de calibration aurait pu être étudiée. Dans nos exemples, les protocoles ont été optimisés selon le critère D pour lequel la preuve de convergence des algorithmes multiplicatifs a été établie, tout comme pour les critères A et c (Yu 2010). Le calcul du critère D_s (obtenu à partir du déterminant d'une partie de la FIM) étant proche de celui du critère D, il pourrait aussi être utilisé lors d'optimisation de protocoles par algorithme multiplicatif. Il a par ailleurs été montré que les algorithmes multiplicatifs étaient compatibles avec l'utilisation de critères d'optimalité combinant deux critères (Holland-Letz 2017). Le critère DD_s , utilisé dans le premier travail (section 2.1.), pourrait donc aussi être utilisé pour trouver les poids optimaux de chaque protocole élémentaire avec l'algorithme multiplicatif. Il serait intéressant d'évaluer la capacité des algorithmes multiplicatifs à optimiser des protocoles selon des critères robustes, utilisés en cas d'incertitude du modèle (par exemple, le critère CD (voir section 2.1.)), des paramètres (par exemple le critère ED) ou des deux (par exemple, le critère CDE). Par ailleurs, il serait intéressant d'étendre l'utilisation de cet algorithme à l'optimisation de variables dans d'autres contextes, par exemple dans les modèles décrivant des données répétées non longitudinales. Nous avons observé que l'utilisation de cet algorithme permet d'obtenir un design pouvant être composé de plusieurs protocoles élémentaires informatifs, ce qui est difficilement réalisable lors d'optimisation combinatoire, comme par exemple dans le premier travail (section 2.1.), où tous les individus étaient affectés à un seul protocole élémentaire. Pour cet exemple commun aux travaux exposés en sections 2.1. et 2.2., l'algorithme multiplicatif a permis de trouver un design D-optimal composé de trois protocoles élémentaires différents. Une des limites du travail présenté est l'absence de comparaison à d'autres algorithmes, par exemple à l'algorithme de Fedorov-Wynn. Des comparaisons pourraient être effectuées à l'avenir, tant en termes de conception, d'utilisation (par exemple, FW nécessite de fournir un protocole initial contrairement aux algorithmes multiplicatifs) que de résultats (protocole optimal et temps de calcul notamment).

Afin d'éviter de passer à côté de protocoles efficaces avec l'algorithme multiplicatif, il est souhaitable, lorsque cela est possible, de définir des contraintes permettant d'évaluer un nombre suffisamment important de protocoles, tout en tenant compte du temps de calcul nécessaire à

l'évaluation de la FIM par protocole élémentaire. Ceci est notamment important pour les méthodes d'évaluation de la FIM pour modèles de données discrètes (telles que QRMC/AGQ ou MC/HMC), beaucoup plus coûteuses en temps de calcul que la méthode FO qui est adaptée aux modèles peu complexes pour données continues. L'algorithme développé dans ce travail requiert une discrétisation de l'espace de design, ce qui permet de trouver le protocole D-optimal dans cet espace défini. L'utilisation d'un algorithme permettant de chercher un design dans un espace continu peut permettre de trouver un protocole plus efficace. Cependant, l'optimisation en temps ou en doses continues peut conduire à des protocoles difficilement réalisables en clinique.

L'algorithme développé dans le cadre de ce travail sera intégré dans la prochaine version de PFIM, PFIM 5.0 (Seurat et al. 2019). Le programme PFIM a été reprogrammé entièrement, une version alpha est actuellement disponible sous forme de librairie R. et la version beta est prévue au cours du premier semestre 2021. Comme indiqué dans la section 2.2.2., la modularité de la programmation orientée-objet permettra d'intégrer de nouvelles fonctions dans le programme à chaque nouvelle version.

Parmi les nouvelles fonctionnalités à intégrer à PFIM 5.0., certaines méthodes ont été développées et ont fait l'objet de publications, mais ne sont pas encore disponibles dans les programmes de planification de protocole. Il existe au moins 3 méthodes permettant d'évaluer adéquatement la FIM pour des modèles de données discrètes : MC/HMC (Riviere et al. 2016), MC/AGQ (Ueckert and Mentré 2017) et QRMC/AGQ (Ueckert and Mentré 2015), il pourrait être intéressant de comparer ces méthodes afin de déterminer la ou lesquelles il serait préférable d'intégrer au programme. Dans la version actuelle, il est uniquement possible d'optimiser un protocole selon le critère D, c'est-à-dire de chercher le protocole permettant d'obtenir les meilleures précisions d'estimations de l'ensemble des paramètres, pour un modèle et ses valeurs de paramètres donnés. Comme expliqué en sections 1.2.1.2, 2.1, et plus haut dans cette section, d'autres critères d'optimalités peuvent avoir leur intérêt pour répondre à d'autres objectifs. Par exemple, le critère DD_s serait intéressant à intégrer au logiciel, pour les cas où l'objectif serait de chercher un protocole permettant une estimation précise de certains paramètres sans sacrifier la précision des autres paramètres. Sont aussi à intégrer dans PFIM les développements méthodologiques concernant les designs robustes, en cas d'incertitude du modèle et/ou des paramètres (Loingeville et al. 2019). Vérifier la convergence des algorithmes vers le protocole optimal selon les différents critères d'optimalité est cependant nécessaire avant la mise à disposition de ces approches.

De nouveaux algorithmes d'optimisation pourraient aussi être proposés dans PFIM, notamment pour l'optimisation dans un espace de design continu. Par exemple, *Particle Swarm Optimization* (PSO), une métaheuristique d'algorithme dont plusieurs applications d'optimisation de protocoles d'études biomédicales ont été effectuées (Qiu et al. 2014). Cette méthode a été étendue et adaptée aux études longitudinales analysées par modèles mixtes, dans le cadre d'une collaboration entre le département de biostatistiques de l'Université de Los Angeles et notre équipe et un article est soumis dans *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. Un autre algorithme, développé dans notre équipe, dérivé de l'algorithme de Metropolis-Hastings et nommé PGBO (*Population Genetics-Based Optimizer*) pourrait aussi être intégré dans PFIM 5.0. (Bauer et al. 2014).

D'autres méthodes en optimisation de protocoles nécessitent d'être développées. Aujourd'hui dans PFIM 4.0., il est possible de prédire la puissance du test de Wald pour des covariables catégorielles à partir de la FIM. Il est envisagé d'étendre cette fonctionnalité aux covariables continues, en supposant une distribution *a priori* de ces covariables au sein de la population. Les logiciels d'estimations de paramètres en pharmacométrie tels que Monolix permettent de définir des distributions logit-normale et probit-normale, il serait intéressant de pouvoir utiliser ces distributions dans PFIM. De même, il serait utile de prendre en compte les données sous la limite de quantification lors de l'évaluation et de l'optimisation de protocoles, la vraisemblance se calculant différemment pour ce type de données (Ahn et al. 2008). L'inverse de la FIM étant la borne inférieure de la matrice de variance-covariance, une pratique courante consiste à utiliser la FIM pour trouver le protocole optimal puis simuler des jeux de données pour vérifier que ce protocole permet réellement de répondre à l'objectif. Il serait intéressant, soit d'intégrer un outil de simulation dans PFIM soit de développer son interopérabilité avec des outils permettant de le faire. Un objectif du programme DDMoRe (<http://www.ddmore.eu/>) est de faciliter le passage entre les différents programmes utilisés en pharmacométrie, d'estimation de paramètres et d'optimisation de protocole. A ce sujet, le programme Pumas vise à intégrer l'ensemble de ces fonctionnalités (<http://dev.pumas.ai/>) et un outil d'évaluation et optimisation de protocoles est aussi disponible dans la version actuelle de NONMEM (NONMEM 7.5.).

4.1.3. Comparaison de protocoles pour évaluer la décroissance tumorale précoce

Pour évaluer un nombre restreint de protocoles, l'approche par simulations est préférable à celle basée sur la FIM, comme expliqué section 1.2.2. Nous avons donc appliqué cette approche en

oncologie, où l'objectif était de comprendre quels protocoles étaient préférables afin de montrer l'efficacité d'une nouvelle combinaison de molécules sur la décroissance tumorale à 8 semaines de traitement (ETS8), par rapport au traitement de référence. Le traitement de référence considéré était le cetuximab (C) à dose standard et la nouvelle combinaison testée était le cetuximab (à dose standard ou inférieure) et une nouvelle molécule M (à dose standard fixée). Nous avons simulé des données de taille tumorale à partir d'un modèle de croissance tumorale incluant l'effet propre de chaque molécule ainsi que l'interaction entre les deux drogues. Pour chaque protocole, trois scénarios ont été étudiés : pas d'effet de M (afin d'investiguer l'erreur de type 1), additivité de C et M et synergie de C et M. Dans ce travail, nous avons montré qu'un protocole à 1 bras, c'est-à-dire la comparaison des résultats du bras expérimental (ici la combinaison des deux anticancéreux à la dose standard) à une valeur de référence obtenue à partir d'une étude précédente, permet de maximiser la puissance statistique pour montrer un effet de la nouvelle thérapie. Cependant, ce type de protocole nécessite de faire l'hypothèse de connaître la valeur à laquelle se comparer pour le traitement de référence. Cela est possible si les conditions (caractéristiques des patients, prise en charge médicale, techniques analytiques, *etc.*) dans lesquelles a été obtenue la valeur de référence sont similaires à celles de la nouvelle étude, or ces conditions peuvent évoluer, en particulier en oncologie. En effet, nous avons montré qu'une erreur sur cette valeur de référence conduit à une inflation de l'erreur de type 1, voire à une réduction de la puissance si la valeur de référence choisie est plus élevée que la vraie valeur. Un protocole incluant un bras avec traitement de référence est donc préférable, ce qui a aussi été montré dans d'autres cadres (Viele et al. 2014). Des protocoles à deux bras (la référence C vs. la combinaison C+M aux doses standards) et quatre bras (C vs. C+M avec différentes doses de C possibles) ont été étudiés. Pour un même nombre total de patients, le protocole à deux bras permettait d'atteindre une meilleure puissance statistique que celui à quatre bras (en cas d'additivité, 69 % vs. 26 %). Un protocole à quatre bras permettait cependant une meilleure exploration de la relation exposition médicamenteuse-effet ainsi qu'un choix de dose de C plus pertinent pour la combinaison C+M. De plus, dans le scénario de synergie entre les deux molécules, la puissance atteinte pour détecter un effet supérieur de la combinaison était supérieure à 60 %, rendant ainsi ce choix de protocole potentiellement pertinent.

Nous avons aussi comparé les résultats obtenus à partir des tests statistiques réalisés avec les valeurs continues à ceux obtenus avec ces valeurs catégorisées en fonction de la valeur de décroissance tumorale à 8 semaines, tel que cela peut être fait pour les critères RECIST (Eisenhauer et al. 2009). Dans notre exemple, la puissance pour détecter un effet de la combinaison dans un protocole à deux bras (essai contrôlé randomisé classique) était de 69 %

en utilisant les valeurs continues, contre 47 % en utilisant une réponse binaire de type répondeur/non répondeur selon la valeur d'ETS8. Cela montre, en accord avec d'autres études, que l'utilisation des critères RECIST n'est pas optimale dans les études cliniques en oncologie (Sharma et al. 2012).

Lors de l'évaluation de protocole par simulations d'essais cliniques, il est recommandé de se rapprocher du contexte réel de l'étude que l'on veut étudier, par exemple en incluant les écarts au protocole théorique (mauvaise adhérence à la thérapie, données manquantes, sorties d'étude, temps de mesure différents de ceux initialement prévus, *etc.*) (Sheiner 2002). Ces éléments n'ont pas été pris en compte dans cette étude de simulation afin de faciliter l'analyse et l'interprétation des résultats, et parce que ces écarts au protocole sont rares dans les 8 premières semaines de thérapie, leur prise en compte aurait donc eu un faible impact sur les résultats.

Ici, les protocoles considérés étaient relativement classiques, mais d'autres possibilités auraient pu être envisagées. Il aurait par exemple été intéressant d'étudier un protocole à 2 bras avec davantage de sujets dans le bras expérimental (C+M) que dans le bras contrôle (C). Un essai adaptatif aurait aussi été intéressant à évaluer dans ce cadre. Lors de la première étape, deux bras seraient considérés : un bras pour la combinaison aux doses standard C+M et un bras contrôle C. Lors d'une analyse intermédiaire, selon un critère défini, par exemple fondé sur une différence d'ETS8 minimale et/ou sur une p-valeur du test de Wilcoxon, l'étude serait arrêtée pour futilité ou continuée, soit avec les deux mêmes bras, soit en ajoutant la possibilité d'étudier la combinaison avec une dose inférieure de C (afin de limiter la toxicité) si l'effet de la combinaison est déjà montré à la fin de la première étape.

De manière générale dans les études cliniques sur une nouvelle combinaison, lorsque cela est possible d'un point de vue éthique, il est préférable de choisir un protocole randomisé à plusieurs bras afin d'identifier au mieux les effets de chaque molécule de la combinaison considérée, notamment lorsque l'étude porte sur une nouvelle combinaison de drogues dont une est le traitement de référence (LoRusso et al. 2010).

4.2. Modélisation de l'interaction bactériophages/bactéries pour le traitement de la pneumonie

Les bactériophages font partie des solutions envisageables pour contrer l'émergence des résistances bactériennes aux antibiotiques. A partir de données issues d'une série d'expériences *in vitro* et chez un modèle murin d'infection pulmonaire, nous avons construit un modèle

permettant notamment de comprendre les interactions phages-bactéries, et de comparer l'efficacité de la phagothérapie selon différentes modalités d'administration. Ce modèle a été construit en plusieurs étapes, à partir de données provenant de différentes expériences. C'est à notre connaissance la première fois que l'on dispose d'une telle richesse de données, en quantité (total de 81 expériences *in vitro* et 241 souris) et en diversité (3 types d'expériences *in vitro* et deux *in vivo*) pour un même couple phage-bactérie afin de comprendre leur interaction à l'aide d'un modèle mathématique.

A partir de données *in vitro*, ont été déterminées des paramètres concernant le cycle d'infection de la bactérie *E.coli* 536 par le bactériophage 536_P1. Par exemple, le *burst size*, c'est-à-dire le nombre de phages libérés à chaque bactérie lysée, a été estimé grâce aux données de *one step growth*. Les paramètres de durée du cycle infectieux (environ 30 minutes) ont été estimés, l'identification de la phase de latence médiane et de la phase d'éclipse médiane pourrait cependant être meilleure grâce à des expériences avec et sans chloroforme, celui-ci permettant de détecter les phages intra-bactériens en plus des phages libres (Rabinovitch et al. 1999). D'autres paramètres ont été estimés à partir des données *in vitro* (expériences de cinétique de lyse et de co-évolution), par exemple le taux d'infectivité (c'est-à-dire la probabilité de rencontre entre phage et bactérie) et les paramètres de résistance (fréquence de fluctuation et valeur sélective). Ces paramètres ont cependant été ré-évalués pour les expériences *in vivo*.

Chez la souris non infectée, la PK des phages a été évaluée à la suite de l'administration IT ou IV d'une dose connue de bactériophages (10^8 PFU). Pour ces expériences, une mesure de la concentration de phage était effectuée uniquement lors du sacrifice des souris. L'analyse de ces données par modèles mixtes fournissait des paramètres incohérents, nous avons donc choisi d'analyser celles-ci par *naïve pooled data*, méthode consistant ici à rassembler toutes les données de souris par voie d'administration et de considérer celles-ci comme un individu. Cette méthode, bien que potentiellement biaisée et moins puissante que les modèles mixtes (Wang et al. 2018), nous a permis d'obtenir des paramètres PK cohérents par rapport à ceux retrouvés dans la littérature et de détecter une différence de vitesse d'élimination entre les deux voies IT et IV, les phages se distribuant probablement davantage plus profondément dans les alvéoles pulmonaires après administration IT (Morello et al. 2011; Oechslin et al. 2017).

Les données issues d'études chez la souris infectée étaient particulièrement innovantes et informatives. Les bactéries étant bioluminescentes, des données longitudinales de charge bactérienne au niveau pulmonaire étaient produites et ont pu largement aider à la description de l'évolution de l'infection. Alors que les MNLEM semblent être l'outil idoine pour décrire les

données issues de cette technologie, ils sont à notre connaissance encore peu utilisés pour ces analyses. Les souris infectées étaient réparties dans 6 groupes de traitement par phage différents : non traitées (contrôle), par voie IV à dose faible (IV Lo) ou intermédiaire (IV Med) ou par voie IT à dose faible (IT Lo), intermédiaire (IT Med) ou haute (IT Hi). Le modèle final décrivant ces données était composé d'équations différentielles ordinaires permettant d'expliquer l'évolution de la charge bactérienne et la cinétique des phages au niveau pulmonaire, en décrivant les interactions entre phages et bactéries. Ce modèle prenait en compte d'autres composantes comme la cinétique propre du phage (telle que décrite chez la souris non infectée) ou le rôle de l'immunité innée dans le contrôle de l'infection. Le modèle a également décrit la probabilité de survie des souris de manière conjointe à la charge bactérienne longitudinale.

On remarque une forte hétérogénéité d'inoculum bactérien réel, alors que celui-ci était théoriquement similaire pour l'ensemble des souris. Cela est probablement lié au mode d'inoculation choisi, par voie intratrachéale, cette manipulation étant plus sujette à une variabilité inter-souris. La variabilité était par exemple moindre lors d'expériences précédentes menées sur le même type de souris, mais par inoculation intranasale (Dufour et al. 2015). Cette variabilité d'inoculum a finalement été estimée au sein du modèle, et a permis d'expliquer des profils très variés d'évolution de charge bactérienne, pour le groupe contrôle comme pour les différents groupes traités par phages. Elle permet aussi d'appréhender la phagothérapie d'un point de vue plus proche d'une réalité clinique où les charges bactériennes sont différentes lors de l'administration.

Lors de la construction du modèle, l'ajout d'un compartiment de bactéries réfractaires aux phages a permis d'assurer une meilleure description des données observées, notamment chez certaines souris présentant un « rebond » de la charge bactérienne après 24 h. Ce phénomène avait aussi été observé *in vitro* lors des expériences de cinétique de lyse et de co-évolution, 10 heures après la mise en contact des phages et des bactéries. Au sein de ces bactéries dites réfractaires, sont probablement regroupées les bactéries résistantes au phages et celles non atteintes par les phages pour des problèmes de répartition spatiale.

On remarque que certains paramètres ont été estimés à des valeurs différentes chez la souris par rapport à celles estimées *in vitro*, notamment le taux d'infectivité β et le taux de lyse bactérienne k . Cela s'explique par la différence de milieu entre ces deux types d'expériences et surtout au rôle important de l'immunité de l'hôte sur l'évolution de la charge bactérienne.

La caractérisation de l'interaction phage-bactérie chez la souris, bien qu'acceptable, pourrait être améliorée par une technologie permettant de suivre le compte pulmonaire de phages de façon longitudinale plutôt qu'au temps de sacrifice uniquement. Le modèle intègre à la fois l'interaction phage-bactérie, mais aussi, de manière relativement simpliste, le rôle de la réponse immunitaire sur la charge bactérienne. Le suivi de marqueurs de l'immunité pourrait aider à caractériser le rôle des défenses de l'hôte de manière plus fine (Roach et al. 2017).

Le modèle développé et les simulations effectuées à partir du modèle permettent d'aider à la compréhension des différences entre les groupes. Pour un inoculum létal (survie de 0 % à 48 h) chez une souris non traitée (correspondant à 10^7 CFU/g), le temps après administration nécessaire à l'obtention d'une concentration de phage supérieure à 9 log est différent selon la voie d'administration IV ou IT et peu différent selon la dose (2,4 h pour IT Hi vs. 5,1 h pour IV Me vs. 6,0 h pour IV Lo). Cela peut conduire à un délai différent pour le contrôle de la charge bactérienne voire à l'impossibilité de réduire cette charge bactérienne durablement, donc à une survie légèrement moins importante à 48 h dans le groupe IV Lo par rapport à IT Hi (66 % vs. 75 %). L'utilisation d'un modèle conjoint, en plus de permettre de réaliser des simulations de survie facilitant l'explication, a l'avantage de corriger les biais potentiels dans l'estimation des paramètres longitudinaux, comme expliqué dans la section 1.1.2.

Le modèle mathématique construit pourrait permettre d'aider au choix de dose et de voie d'administration pour la planification de protocoles de futures études de phagothérapie, et à terme pour les cliniciens. La structure du modèle mathématique construit servira aussi à la description de futures données issues d'expérimentations similaires. Les perspectives en termes d'expérimentation animale en phagothérapie sont d'étudier l'évolution de la charge bactérienne chez l'animal traité par un cocktail de plusieurs phages, dans le but de limiter le phénomène de résistance aux phages et donc d'accroître l'efficacité de ce traitement. L'ensemble de ces études permettront d'optimiser la prise en charge de pneumonies bactériennes par les bactériophages en clinique, en termes de choix du phage (ou des phages pour un cocktail) et du schéma d'administration.

4.3. Conclusion générale

Ces travaux de doctorat sont inscrits dans la continuité des développements méthodologiques autour de la FIM. Nous avons dans ce cadre développé une approche de protocole robuste en cas d'incertitude du modèle et un algorithme pour l'optimisation de protocole d'études analysées par MNLEM. Ces outils serviront donc pour la planification de protocoles d'études

longitudinales, l'algorithme développé étant intégré à la nouvelle version du logiciel PFIM, dont la version alpha est disponible. L'utilisation de l'approche de protocole robuste pourra aussi être facilitée lorsqu'elle sera intégrée dans PFIM. Une application d'évaluation de protocoles en phase précoce d'essai sur une combinaison d'anticancéreux a aussi été effectuée, de manière à montrer les avantages et inconvénients de différents designs et ainsi aider aux choix des cliniciens dans ce cadre. Nous avons utilisé les MNLEM pour décrire l'interaction entre les phages et les bactéries. Le modèle construit aidera à planifier les prochaines études en phagothérapie. Certaines expériences pourront d'ailleurs être planifiées à l'aide de PFIM, éventuellement en utilisant la méthodologie des protocoles robustes, car une incertitude sur les modèles et les valeurs de paramètres pourrait être importante pour de nouveaux couples phages/bactéries et de nouveaux modèles animaux.

Nous espérons avoir démontré l'intérêt de la planification de protocoles, en utilisant la FIM ou les simulations, ainsi que l'intérêt des MNLEM dans les études longitudinales afin de contribuer au progrès en santé publique.

Bibliographie

Abedon ST, Kuhl SJ, Blasdel BG, Kutter EM. Phage treatment of human infections. *Bacteriophage*. 2011;1(2):66–85.

Ahn JE, Karlsson MO, Dunne A, Ludden TM. Likelihood based approaches to handling data below the quantification limit using NONMEM VI. *J Pharmacokinet Pharmacodyn*. 2008;35(4):401–21.

Aliev A, Fedorov V, Leonov S, McHugh B, Magee M. PkStaMp library for constructing optimal population designs for PK/PD studies. *Commun Stat Simul Comput*. 2012;41(6):717–29.

Atkinson A, Donev A, Tobias R. *Optimum experimental designs, with SAS*. Oxford, New York: Oxford University Press; 2007.

Atkinson AC, Bogacka B. Compound D- and Ds-optimum designs for determining the order of a chemical reaction. *Technometrics*. 1997;39(4):347–56.

al-Banna MK, Kelman AW, Whiting B. Experimental design and efficient parameter estimation in population pharmacokinetics. *J Pharmacokinet Biopharm*. 1990;18(4):347–60.

Bauer R, Mentré F, Kaddouri H, Le Bras J, Le Nagard H. Benefits of a new Metropolis–Hasting based algorithm, in non-linear regression for estimation of ex vivo antimalarial sensitivity in patients infected with two strains. *Comput Biol Med*. 2014;55:16–25.

Bauer RJ. NONMEM tutorial part I: description of commands and options, with simple examples of population analysis. *CPT Pharmacometrics Syst Pharmacol*. 2019a;8(8):525–37.

Bauer RJ. NONMEM tutorial part II: estimation methods and advanced examples. *CPT Pharmacometrics Syst Pharmacol*. 2019b;(8):538–56.

Berthaud R, Benaboud S, Hirt D, Genuini M, Oualha M, Castelle M, et al. Early Bayesian dose adjustment of vancomycin continuous infusion in children: a randomized controlled trial. *Antimicrob Agents Chemother*. 2019;63(12):e01102-19.

Bhandare S, Colom J, Baig A, Ritchie JM, Bukhari H, Shah MA, et al. Reviving phage therapy for the treatment of cholera. *J Infect Dis*. 2019;219(5):786–94.

Björnsson MA, Simonsson USH. Modelling of pain intensity and informative dropout in a dental pain model after naproxen, naproxen and placebo administration. *Br J Clin Pharmacol*. 2011;71(6):899–906.

Blain H, Rolland Y, Tuailon E, Giacosa N, Albrand M, Jaussent A, et al. Efficacy of a test-retest strategy in residents and health care personnel of a nursing home facing a COVID-19 outbreak. *J Am Med Dir Assoc*. 2020;21(7):933–6.

Bonate PL. Clinical trial simulation in drug development. *Pharm Res*. 2000;17(3):252–6.

Brendel K, Comets E, Laffont C, Laveille C, Mentré F. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. *Pharm Res*. 2006;23(9):2036–49.

Buatois S, Ueckert S, Frey N, Retout S, Mentré F. Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. *AAPS J*. 2018;20(3):56.

Buckland ST, Burnham KP, Augustin NH. Model selection: an integral part of inference. *Biometrics*. 1997;53(2):603–18.

Cafilisch KM, Suh GA, Patel R. Biological challenges of phage therapy and proposed solutions: a literature review. *Expert Rev Anti Infect Ther*. 2019;17(12):1011–41.

Cairns BJ, Timms AR, Jansen VAA, Connerton IF, Payne RJH. Quantitative models of in vitro bacteriophage-host dynamics and their application to phage therapy. *PLoS Pathog*. 2009;5(1):e1000253.

Cerou M, Lavielle M, Brendel K, Chenel M, Comets E. Development and performance of npde for the evaluation of time-to-event models. *Pharm Res*. 2018;35(2):30.

Chatelut E, Canal P, Brunner V, Chevreau C, Pujol A, Boneu A, et al. Prediction of carboplatin clearance from standard morphological and biological patient characteristics. *J Natl Cancer Inst*. 1995;87(8):573–80.

Claret L, Girard P, Hoff PM, Van Cutsem E, Zuideveld KP, Jorga K, et al. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol*. 2009;27(25):4103–8.

Combes FP, Retout S, Frey N, Mentré F. Prediction of shrinkage of individual parameters using the bayesian information matrix in non-linear mixed effect models with evaluation in pharmacokinetics. *Pharm Res*. 2013;30(9):2355–67.

Comets E, Lavenu A, Lavielle M. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. *J Stat Softw*. 2017;80(3).

Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol*. 1972;34(2):187–220.

- Davidian M, Giltinan DM. Nonlinear Models for Repeated Measurement Data. 1st ed. Routledge; 2017.
- Delattre M, Lavielle M, Poursat M-A. A note on BIC in mixed-effects models. *Electron J Stat.* 2014;8(1):456–75.
- Desmée S, Mentré F, Veyrat-Follet C, Guedj J. Nonlinear mixed-effect models for prostate-specific antigen kinetics and link with survival in the context of metastatic prostate cancer: a comparison by simulation of two-stage and joint approaches. *AAPS J.* 2015;17(3):691–9.
- Duffull S, Waterhouse T, Eccleston J. Some considerations on the design of population pharmacokinetic studies. *J Pharmacokinet Pharmacodyn.* 2005;32(3–4):441–57.
- Duffull SB, Retout S, Mentré F. The use of simulated annealing for finding optimal population designs. *Comput Methods Programs Biomed.* 2002;69(1):25–35.
- Dufour N, Debarbieux L, Fromentin M, Ricard J-D. Treatment of highly virulent extraintestinal pathogenic *Escherichia coli* pneumonia with bacteriophages. *Crit Care Med.* 2015;43(6):e190–198.
- Dufour N, Delattre R, Chevallereau A, Ricard J-D, Debarbieux L. Phage therapy of pneumonia is not associated with an overstimulation of the inflammatory response compared to antibiotic treatment in mice. *Antimicrob Agents Chemother.* 2019;63(8).
- Dufour N, Delattre R, Ricard J-D, Debarbieux L. The lysis of pathogenic *Escherichia coli* by bacteriophages releases less endotoxin than by β -lactams. *Clin Infect Dis.* 2017;64(11):1582–8.
- Dumont C, Lestini G, Le Nagard H, Mentré F, Comets E, Nguyen TT, et al. PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models. *Comput Methods Programs Biomed.* 2018;156:217–29.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45(2):228–47.
- Ette EI, Williams PJ. *Pharmacometrics: the science of quantitative pharmacology.* Hoboken, NJ: Wiley; 2007.
- Faruque SM, Islam MJ, Ahmad QS, Faruque ASG, Sack DA, Nair GB, et al. Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *Proc Natl Acad Sci USA.* 2005;102(17):6119–24.
- Fedorov V. *Theory of optimal experiments.* New York: Acad. Press; 1972.

Fedorov V, Leonov S. Optimal design for nonlinear response models. CRC Press. 2013.

Fisher R. On the mathematical foundations of theoretical statistics. *Phil Trans R Soc A Philos Trans R Soc A-Math Phys Eng Sci.* 1922;222(594–604):309–68.

Foo L-K, Duffull S. Methods of robust design of nonlinear models with an application to pharmacokinetics. *J Biopharm Stat.* 2010;20(4):886–902.

Food and Drug Administration. Population pharmacokinetics: guidance for industry. 2019.

Foracchia M, Hooker A, Vicini P, Ruggeri A. POPED, a software for optimal experiment design in population kinetics. *Comput Methods Programs Biomed.* 2004;74(1):29–46.

Fragkou PC, Belhadi D, Peiffer-Smadja N, Moschopoulos CD, Lescure F-X, Janocha H, et al. Review of trials currently testing treatment and prevention of COVID-19. *Clin Microbiol Infect.* 2020;26(8):988–98.

Friberg L, de Greef R, Kerbusch T, Karlsson M. Modeling and simulation of the time course of asenapine exposure response and dropout patterns in acute schizophrenia. *Clin Pharmacol Ther.* 2009;86(1):84–91.

Friberg LE, Guedj J. Acute bacterial or viral infection-What's the difference? A perspective from PKPD modellers. *Clin Microbiol Infect.* 2019;Ahead of print.

Furfaro LL, Payne MS, Chang BJ. Bacteriophage therapy: clinical trials and regulatory hurdles. *Front Cell Infect Microbiol.* 2018;8(376).

Gompertz B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos Trans R Soc Lond.* 1825;115:513–83.

Górski A, Międzybrodzki R, Łobočka M, Głowacka-Rutkowska A, Bednarek A, Borysowski J, et al. Phage therapy: what have we learned? *Viruses.* 2018;10(6).

Green B, Duffull SB. Development of a dosing strategy for enoxaparin in obese patients: A dosing strategy for enoxaparin in obese patients. *Br J Clin Pharmacol.* 2003;56(1):96–103.

Gu J, Liu X, Li Y, Han W, Lei L, Yang Y, et al. A method for generation phage cocktail with great therapeutic potential. *PLoS ONE.* 2012;7(3):e31698.

Gueorguieva I, Ogungbenro K, Graham G, Glatt S, Aarons L. A program for individual and population optimal design for univariate and multivariate response pharmacokinetic–pharmacodynamic models. *Comput Methods Programs Biomed.* 2007;86(1):51–61.

Guo B, Abdelraouf K, Ledesma KR, Chang K-T, Nikolaou M, Tam VH. Quantitative impact of neutrophils on bacterial clearance in a murine pneumonia model. *Antimicrob Agents Chemother.* 2011;55(10):4601–5.

Hankin M. L'action bactéricide des eaux de la Jumna et du Gange sur le vibrion du choléra. *Ann Inst Pasteur.* 1896;(10):511–23.

d'Herelle F. Sur un microbe invisible antagoniste des bacilles dysentériques. *C R Acad Sci Paris.* 1917;(165):373–5.

d'Herelle F. Bacteriophage as a treatment in acute medical and surgical infections. *Bull N Y Acad Med.* 1931;7(5):329–48.

Holland-Letz T. On the combination of c- and D-optimal designs: general approaches and applications in dose-response studies. *Biometrics.* 2017;73(1):206–13.

Jault P, Leclerc T, Jennes S, Pirnay JP, Que Y-A, Resch G, et al. Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by *Pseudomonas aeruginosa* (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial. *Lancet Infect Dis.* 2019;19(1):35–45.

Jensen EC, Schrader HS, Rieland B, Thompson TL, Lee KW, Nickerson KW, et al. Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. *Appl Environ Microbiol.* 1998;64(2):575–80.

Jonsson EN, Wade JR, Karlsson MO. Comparison of some practical sampling strategies for population pharmacokinetic studies. *J Pharmacokinet Pharmacodyn.* 1996;24(2):245–63.

Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53(282):457.

Karlin S, Studden WJ. Optimal experimental designs. *Ann Math Stat.* 1966;37(4):783–815.

Karlsson MO, Sheiner LB. The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J Pharmacokinet Pharmacodyn.* 1993;21(6):735–50.

Kiefer J. Optimum experimental designs. *J R Stat Soc Series B Stat Methodol.* 1959;21(2):272–319.

Kimko H, Pinheiro J. Model-based clinical drug development in the past, present and future: a commentary. *Br J Clin Pharmacol.* 2015;79(1):108–16.

Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal.* 2005;49(4):1020–38.

- Kutateladze M, Adamia R. Phage therapy experience at the Eliava Institute. *Med Mal Infect.* 2008;38(8):426–30.
- Landerman LR, Mustillo SA, Land KC. Modeling repeated measures of dichotomous data: testing whether the within-person trajectory of change varies across levels of between-person factors. *Soc Sci Res.* 2011;40(5):1456–64.
- Lestini G, Dumont C, Mentré F. Influence of the size of cohorts in adaptive design for nonlinear mixed effects models: an evaluation by simulation for a pharmacokinetic and pharmacodynamic model for a biomarker in oncology. *Pharm Res.* 2015;32(10):3159–69.
- Letarov A, Kulikov E. The bacteriophages in human- and animal body-associated microbial communities. *J Appl Microbiol.* 2009;107(1):1–13.
- Levin BR, Bull JJ. Population and evolutionary dynamics of phage therapy. *Nat Rev Microbiol.* 2004;2(2):166–73.
- Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics.* 1990;46(3):673.
- Liu D, Zhang H. Residuals and diagnostics for ordinal regression models: a surrogate approach. *J Am Stat Assoc.* 2018;113(522):845–54.
- Loingeville F, Nguyen TT, Riviere M-K, Mentré F. Robust designs in longitudinal studies accounting for parameter and model uncertainties - application to count data. *J Biopharm Stat.* 2019;1–15.
- LoRusso PM, Boerner SA, Seymour L. An overview of the optimal planning, design, and conduct of phase I studies of new therapeutics. *Clin Cancer Res.* 2010;16(6):1710–8.
- Lu J-F, Claret L, Sutjandra L, Kuchimanchi M, Melara R, Bruno R, et al. Population pharmacokinetic/pharmacodynamic modeling for the time course of tumor shrinkage by motesanib in thyroid cancer patients. *Cancer Chemother Pharmacol.* 2010;66(6):1151–8.
- Madelain V, Baize S, Jacquot F, Reynard S, Fizet A, Barron S, et al. Ebola viral dynamics in nonhuman primates provides insights into virus immuno-pathogenesis and antiviral strategies. *Nat Commun.* 2018;9(1):4013.
- Madelain V, Guedj J, Mentré F, Nguyen THT, Jacquot F, Oestereich L, et al. Favipiravir pharmacokinetics in nonhuman primates and insights for future efficacy studies of hemorrhagic fever viruses. *Antimicrob Agents Chemother.* 2017;61(1).

Marshall S, Madabushi R, Manolis E, Krudys K, Staab A, Dykstra K, et al. Model-informed drug discovery and development: current industry good practice and regulatory expectations and future perspectives. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(2):87–96.

Matsuda T, Freeman TA, Hilbert DW, Duff M, Fuortes M, Stapleton PP, et al. Lysis-deficient bacteriophage therapy decreases endotoxin and inflammatory mediator release and improves survival in a murine peritonitis model. *Surgery*. 2005;137(6):639–46.

McEwen SA, Collignon PJ. Antimicrobial resistance: a one health perspective. *Microbiol Spectr*. 2018;6(2):521–47.

Mentré F, Burtin P, Merlé Y, van Bree J, Mallet A, Steimer J-L. Sparse-sampling optimal designs in pharmacokinetics and toxicokinetics. *Drug Inf J*. 1995;29(3):997–1019.

Mentré F, Mallet A, Baccar D. Optimal design in random-effects regression models. *Biometrika*. 1997;84(2):429–42.

Miller P. Role of pharmacoeconomic analysis in R&D decision making: when, where, how? *Pharmacoeconomics*. 2005;23(1):1–12.

Morello E, Saussereau E, Maura D, Huerre M, Touqui L, Debarbieux L. Pulmonary bacteriophage therapy on *Pseudomonas aeruginosa* cystic fibrosis strains: first steps towards treatment and prevention. *PLoS ONE*. 2011;6(2):e16963.

Mouton JW, Vinks AA, Punt NC. Pharmacokinetic-pharmacodynamic modeling of activity of ceftazidime during continuous and intermittent infusion. *Antimicrob Agents Chemother*. 1997;41(4):733–8.

Nash JC. The (Dantzig) simplex method for linear programming. *Comput Sci Eng*. 2000;2(1):29–31.

Nelder JA, Mead R. A simplex method for function minimization. *Comput J*. 1965;7(4):308–13.

Nguyen THT, Mouksassi M-S, Holford N, Al-Huniti N, Freedman I, Hooker AC, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT Pharmacometrics Syst Pharmacol*. 2017;6(2):87–109.

Nguyen TT, Bazzoli C, Mentré F. Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models. *Stat Med*. 2012;31(11–12):1043–58.

Nguyen TT, Bénech H, Delaforge M, Lenuzza N. Design optimisation for pharmacokinetic modeling of a cocktail of phenotyping drugs. *Pharm Stat*. 2016;15(2):165–77.

Nguyen TT, Mentré F. Evaluation of the Fisher information matrix in nonlinear mixed effect models using adaptive Gaussian quadrature. *Comput Stat Data Anal.* 2014;80:57–69.

Nielsen EI, Friberg LE. Pharmacokinetic-pharmacodynamic modeling of antibacterial drugs. Andersson D, editor. *Pharmacol Rev.* 2013;65(3):1053–90.

Nielsen EI, Viberg A, Löwdin E, Cars O, Karlsson MO, Sandström M. Semimechanistic pharmacokinetic/pharmacodynamic model for assessment of activity of antibacterial agents from time-kill curve experiments. *Antimicrob Agents Chemother.* 2007;51(1):128–36.

Nyberg J, Bazzoli C, Ogungbenro K, Aliev A, Leonov S, Duffull S, et al. Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies. *Br J Clin Pharmacol.* 2015;79(1):6–17.

Ochs HD, Buckley RH, Kobayashi RH, Kobayashi AL, Sorensen RU, Douglas SD, et al. Antibody responses to bacteriophage phi X174 in patients with adenosine deaminase deficiency. *Blood.* 1992;80(5):1163–71.

Oechslin F, Piccardi P, Mancini S, Gabard J, Moreillon P, Entenza JM, et al. Synergistic interaction between phage therapy and antibiotics clears *Pseudomonas aeruginosa* infection in endocarditis and reduces virulence. *J Infect Dis.* 2017;215(5):703–12.

O’Neill J. The review on antimicrobial resistance. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. [Internet]. 2014 Dec. Available from: http://www.jpiaamr.eu/wp-content/uploads/2014/12/AMR-Review-Paper-Tackling-a-crisis-for-the-health-and-wealth-of-nations_1-2.pdf

Papathanasiou T, Strathe A, Overgaard RV, Lund TM, Hooker AC. Optimizing dose-finding studies for drug combinations based on exposure-response models. *AAPS J.* 2019;21(5):95.

Patey O, McCallin S, Mazure H, Liddle M, Smithyman A, Dublanchet A. Clinical indications and compassionate use of phage therapy: personal experience and literature review with a focus on osteoarticular infections. *Viruses.* 2018;11(1):18.

Payne R. Phage therapy: The peculiar kinetics of self-replicating pharmaceuticals. *Clin Pharmacol Ther.* 2000;68(3):225–30.

Payne RJ, Jansen VA. Understanding bacteriophage therapy as a density-dependent kinetic process. *J Theor Biol.* 2001;208(1):37–48.

Perelson AS. Modelling viral and immune system dynamics. *Nat Rev Immunol.* 2002;2(1):28–36.

Pinheiro JC, Bates DM. *Mixed-Effects Models in Sand S-PLUS.* Springer New York; 2000.

Pronzato L, Pázman A. Design of experiments in nonlinear models: asymptotic normality, optimality criteria and small-sample properties. Springer Science & Business Media; 2013.

Pronzato L, Walter E. Robust experiment design via stochastic approximation. *Math Biosci.* 1985;75(1):103–20.

Qiu J, Chen R-B, Wang W, Wong WK. Using animal instincts to design efficient biomedical studies via particle swarm optimization. *Swarm Evol Comput.* 2014;18:1–10.

Rabinovitch A, Hadas H, Einav M, Melamed Z, Zaritsky A. Model for bacteriophage T4 development in *Escherichia coli*. *J Bacteriol.* 1999;181(5):1677–83.

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.* 2007;26(28):5162–79.

Retout S, Duffull S, Mentré F. Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. *Comput Methods Programs Biomed.* 2001;65(2):141–51.

Retout S, Mentré F, Bruno R. Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Stat Med.* 2002;21(18):2623–39.

Ribba B, Holford NH, Magni P, Trocóniz I, Gueorguieva I, Girard P, et al. A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT Pharmacometrics Syst Pharmacol.* 2014;3:e113.

Riglet F, Mentré F, Veyrat-Follet C, Bertrand J. Bayesian individual dynamic predictions with uncertainty of longitudinal biomarkers and risks of survival events in a joint modelling framework: a comparison between Stan, Monolix, and NONMEM. *AAPS J.* 2020;22(2):50.

Riviere M-K, Mentré F. R package MIXFIM, version 1.0 <http://mc-stan.org/>. 2015.

Riviere M-K, Ueckert S, Mentré F. An MCMC method for the evaluation of the Fisher information matrix for non-linear mixed effect models. *Biostatistics.* 2016;17(4):737–50.

Rizopoulos D. Joint models for longitudinal and time-to-event data: with applications in R. 0 ed. Chapman and Hall/CRC; 2012.

Roach DR, Leung CY, Henry M, Morello E, Singh D, Di Santo JP, et al. Synergy between the host immune system and bacteriophage is essential for successful phage therapy against an acute respiratory pathogen. *Cell Host Microbe.* 2017;22(1):38-47.e4.

Rodger AJ, Cambiano V, Bruun T, Vernazza P, Collins S, Degen O, et al. Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. *The Lancet*. 2019;393(10189):2428–38.

Rohwer F. Global phage diversity. *Cell*. 2003;113(2):141.

Rosenfeld PJ, Brown DM, Heier JS, Boyer DS, Kaiser PK, Chung CY, et al. Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2006;355(14):1419–31.

Rubinstein A, Mizrahi Y, Bernstein L, Shliozberg J, Golodner M, Liu GQ, et al. Progressive specific immune attrition after primary, secondary and tertiary immunizations with bacteriophage phi X174 in asymptomatic HIV-1 infected patients. *AIDS*. 2000;14(4):F55-62.

Sarker SA, Sultana S, Reuteler G, Moine D, Descombes P, Charton F, et al. Oral phage therapy of acute bacterial diarrhea with two coliphage preparations: a randomized trial in children from Bangladesh. *EBioMedicine*. 2016;4:124–37.

Seurat J, Tang Y, Nguyen TT, Mentré F, Le Nagard H, on behalf of the PFIM group. New version of PFIM for optimal design in nonlinear mixed effects models using R S4. PAGE, Stockholm, Sweden; 2019.

Seymour LK, Calvert AH, Lobbezoo MW, Eisenhauer EA, Giaccone G, task force on methodology for the development of innovative cancer therapies. Design and conduct of early clinical studies of two or more targeted anticancer therapies: recommendations from the task force on methodology for the development of innovative cancer therapies. *Eur J Cancer*. 2013;49(8):1808–14.

Sharma MR, Maitland ML, Ratain MJ. RECIST: no longer the sharpest tool in the oncology clinical trials toolbox. *Cancer Res*. 2012;72(20):5145–50.

Sheiner LB. Is intent-to-treat analysis always (ever) enough? *Br J Clin Pharmacol*. 2002;54(2):203–11.

Sheiner LB, Rosenberg B, Melmon KL. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput Biomed Res*. 1972;5(5):411–59.

Simeoni M, Magni P, Cammia C, De Nicolao G, Croci V, Pesenti E, et al. Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Research*. 2004;64(3):1094–101.

Skurnik M, Strauch E. Phage therapy: facts and fiction. *Int J Med Microbiol*. 2006;296(1):5–14.

Smith AM, Perelson AS. Influenza A virus infection kinetics: quantitative data and models: Modeling influenza kinetics. *Wiley Interdiscip Rev Syst Biol Med*. 2011;3(4):429–45.

Smoragiewicz M, Adjei AA, Calvo E, Tabernero J, Marabelle A, Massard C, et al. Design and conduct of early clinical studies of immunotherapy: recommendations from the task force on Methodology for the Development of Innovative Cancer Therapies 2019 (MDICT). *Clinical Cancer Research*. 2020;26(11):2461–5.

Sölkner J. Choice of optimality criteria for the design of crossbreeding experiments. *J Anim Sci*. 1993;71(11):2867–73.

Stan Development Team. Rstan: the R interface to stan, version 2.12.0. <http://mc-stan.org/>. 2016.

Sulakvelidze A, Morris JG. Bacteriophages as therapeutic agents. *Ann Med*. 2001;33(8):507–9.

Sullivan DC, Schwartz LH, Zhao B. The imaging viewpoint: how imaging affects determination of progression-free survival. *Clin Cancer Res*. 2013;19(10):2621–8.

Syrjänen RK, Kilpi TM, Kaijalainen TH, Herva EE, Takala AK. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Finnish children younger than 2 years old. *J Infect Dis*. 2001;184(4):451–9.

Tardivon C, Desmée S, Kerioui M, Bruno R, Wu B, Mentré F, et al. Association between tumor size kinetics and survival in patients with urothelial carcinoma treated with atezolizumab: implication for patient follow-up. *Clin Pharmacol Ther*. 2019;106(4):810–20.

Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc*. 1995;90(429):27.

Tsulukidze A. Experience of use of bacteriophages in the conditions of war traumatism (in Russian). Tbilisi : Gruzmedgiz. 1941.

Ueckert S, Mentré F. Computation of the Fisher information matrix for discrete nonlinear mixed effect models. 8th International Conference on Computational and Methodological Statistics, London, UK; 2015.

Ueckert S, Mentré F. A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature. *Comput Stat Data Anal*. 2017;111:203–19.

Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat.* 2014;13(1):41–54.

Wählby U, Thomson AH, Milligan PA, Karlsson MO. Models for time-varying covariates in population pharmacokinetic-pharmacodynamic analysis. *Br J Clin Pharmacol.* 2004;58(4):367–77.

Waite TW, Woods DC. Designs for generalized linear models with random block effects via information matrix approximations. *Biometrika.* 2015;102(3):677–93.

Wang S, Zhang Y, Hooker A, Karlsson MO. A comparison between nonlinear mixed effects and naïve pooled data methods in population PK model selection. PAGE, Montreux, Suisse; 2018.

Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, et al. Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clin Pharmacol Ther.* 2009;86(2):167–74.

Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, et al. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature.* 1995;373(6510):117–22.

Weinbauer MG. Ecology of prokaryotic viruses. *FEMS Microbiol Rev.* 2004;28(2):127–81.

Weld RJ, Butts C, Heinemann JA. Models of phage growth and their applicability to phage therapy. *J Theor Biol.* 2004;227(1):1–11.

Wicha SG, Chen C, Clewe O, Simonsson USH. A general pharmacodynamic interaction model identifies perpetrators and victims in drug interactions. *Nat Commun.* 2017;8(1):2129.

Witelski T, Bowen M. *Methods of mathematical modelling: continuous systems and differential equations.* Cham Heidelberg New York Dordrecht London: Springer; 2015.

Wright A, Hawkins CH, Anggård EE, Harper DR. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. *Clin Otolaryngol.* 2009;34(4):349–57.

Wu L, Liu W, Yi GY, Huang Y. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Electron J Stat.* 2012;2012:1–17.

Wunderink RG, Waterer GW. Community-acquired pneumonia. Solomon CG, editor. *N Engl J Med.* 2014;370(6):543–51.

Wynn HP. Results in the theory and construction of D-optimum experimental designs. *J R Stat Soc Series B Stat Methodol.* 1972;34(2):133–47.

Yano Y, Beal SL, Sheiner LB. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *J Pharmacokinet Pharmacodyn*. 2001;28(2):171–92.

Yu J, Wang N, Kågedal M. A new method to model and predict progression free survival based on tumor growth dynamics. *CPT Pharmacometrics Syst Pharmacol*. 2020;9(3):177–84.

Yu Y. Monotonic convergence of a general algorithm for computing optimal designs. *Ann Statist*. 2010;38(3):1593–606.

Annexe

New version of PFIM for optimal design in nonlinear mixed effects models using R S4.
Poster présenté au *Population Approach Group Europe (PAGE)*. Stockholm, Suède. 2019.



New version of PFIM for optimal design in nonlinear mixed effects models using R S4

Jérémy Seurat, Yuxin Tang, Thu Thuy Nguyen, France Menétré, Hervé le Nagard, on behalf of PFIM group

IAME, UMR 1137, INSERM, Université Paris Diderot, Paris, France



Introduction

Background

- Nonlinear mixed effect models are increasingly used for the analysis of longitudinal studies in drug development. Using the Fisher Information Matrix (FIM) to optimize the design of these studies is an efficient alternative to clinical trial simulation
- PFIM 4.0 is a R program devoted to the design evaluation and optimisation [1]
- S4 is an object-oriented programming language which can offer several advantages compared to the traditional programming [2]

Objectives

- To increase the clarity of the different defined elements in PFIM
- To increase the modularity of PFIM, resulting in simplifying the use (manipulation of the different elements) and in facilitating implementation of new functions

The new PFIM

Methods

- As in PFIM 4.0., the model and its parameter values should be given to evaluate the FIM for a given design, or to optimise the design (Figure 1)
- The FIM is evaluated by first order linearisation of the model [3]
- Following an object-oriented approach, PFIM is implemented in R S4 language. The new PFIM can be represented as a class diagram, with PFIM as a central main object (Figure 2)

Implemented features (New features)

- Model
 - Analytic and Ordinary Differential Equation (ODE) form with one or several responses (e.g. PKPD model)
 - Error models: Constant, Proportional, Combined
 - Model parameters: possible different distributions (normal and log-normal) in the same project
- Design evaluation: Expected standard errors of parameters, FIM determinant, eigenvalues
- Design optimisation given constraints: **Multiplicative algorithm** [4], with possibility to optimise both sampling times and doses based on D-criterion

PFIM use

- The following objects have to be defined in a PFIM project:
 - The project name
 - The model
 - Equations of each compartment/response
 - Parameters: fixed effects (μ), standard deviation of random effects (ω) and distribution
 - The design: number of subjects, sampling times, doses, constraints in case of optimisation
- FIM computation & design evaluation/optimisation are performed
- Results can be presented in a summary
- All the elements (model, design and results) can be saved/reloaded/manipulated in R.



Figure 1: FIM-based optimal design

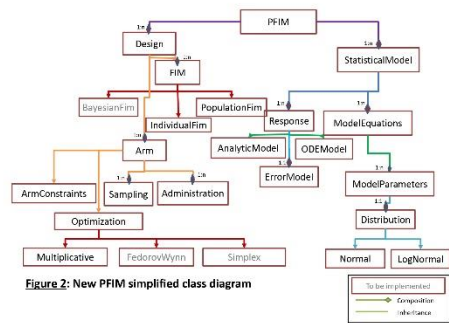


Figure 2: New PFIM simplified class diagram

Example of design evaluation for a PKPD model with the new PFIM

```
## Example of design evaluation in new PFIM
## Create a project
myproj <- newProject("myproj")
myproj@name <- "myproj"
myproj@description <- "Example of design evaluation in new PFIM"
myproj@save()

## Create a model
myproj@model <- newModel("myproj", "myproj")
myproj@model@name <- "myproj"
myproj@model@description <- "Example of design evaluation in new PFIM"
myproj@model@save()

## Create a design
myproj@design <- newDesign("myproj", "myproj")
myproj@design@name <- "myproj"
myproj@design@description <- "Example of design evaluation in new PFIM"
myproj@design@save()

## Evaluate the FIM
myproj@evaluateFIM()

## Print the results
myproj@print()

## Save the results
myproj@save()
```

Figure 3: Example of design evaluation in new PFIM

- Example description
 - PKPD Model written as ODE (Parameters are given in Table 1)
 - Monocompartmental PK
 - Turnover/full Imax PKPD
- Design
 - N=32
 - Oral administration, dose = 100 mg
 - 12 PK samples from 0.5 to 120 hr, 8 PD samples from 0 to 144 hr
- The complete R script is given in Figure 3

Table 1: Model parameters

Parameter	Value	SE	Distribution	Unit
ka (h ⁻¹)	1.9	0.98	lognormal	-
V (L)	8	0.13	lognormal	-
Cl (L/h)	0.13	0.28	lognormal	-
Imax (mg)	5.4	0.45	lognormal	-
Emax (h ⁻¹)	0.06	0.13	lognormal	-
CO ₂ (L/h)	1.2	0.1	lognormal	-
sigma _{PK}	-	-	-	0.8
sigma _{PD}	-	-	-	0.07

μ: fixed effect
σ: standard deviation of random effect
σ_{res}: standard deviation of the residual error

```
## Expected standard errors with new PFIM (left) and PFIM 4.0. (right)
## Expected standard errors
**** Fixed effects ****
value      se      pval
ka      1.90 0.0989929  32.24447
V      8.00 0.134970219  2.71993
Cl      0.13 0.025703877  4.70460
Imax    5.40 0.44401357  0.298877
Emax    0.06 0.020288822  2.34572
CO2     1.20 0.091675680  1.77040
**** Variance of random effects ****
sigma_ka  0.78 0.14713116  26.76020
sigma_V   0.12 0.023593462  29.79132
sigma_Cl  0.28 0.02608758  16.82336
sigma_Imax 0.20 0.02302777  26.59189
sigma_Emax 0.22 0.02608758  22.48244
sigma_CO2 0.21 0.02472328  104.72338
**** Error model ****
sigma_resPK 0.60 0.0861762  14.52516
sigma_resPD 0.07 0.01256809  1.92428
sigma_resPD2 4.90 0.2202782  1.60297
**** Expected standard errors ****
**** Fixed effects ****
value      se      pval
ka      1.90 0.115395815  15.37018
V      8.00 0.134970207  2.71993
Cl      0.13 0.025703877  4.70460
Imax    5.40 0.44401357  0.298877
Emax    0.06 0.020288822  2.34572
CO2     1.20 0.091675680  1.77040
**** Variance of inner subject random effects ****
sigma_ka  0.78 0.14713116  26.76020
sigma_V   0.12 0.023593462  29.79132
sigma_Cl  0.28 0.02608758  16.82336
sigma_Imax 0.20 0.02302777  26.59189
sigma_Emax 0.22 0.02608758  22.48244
sigma_CO2 0.21 0.02472328  104.72338
**** Error model ****
sigma_resPK 0.60 0.0861762  14.52516
sigma_resPD 0.07 0.01256809  1.92428
sigma_resPD2 4.90 0.2202782  1.60297
**** Standard deviation of residual error ****
sigma_resPK 0.60 0.0861762  14.52516
sigma_resPD 0.07 0.01256809  1.92428
sigma_resPD2 4.90 0.2202782  1.60297
```

Figure 4: Expected standard errors with new PFIM (left) and PFIM 4.0. (right)

- The expected standard errors of parameters are consistent with those obtained with PFIM 4.0 (Figure 4)

Perspectives

- To implement features which are in PFIM 4.0.: Fedorov-Wynn and Simplex algorithms for design optimisation, Bayesian FIM to give shrinkage predictions [5], discrete covariates and Wald test power predictions [6]
- To implement new features such as alternative methods to evaluate the FIM (e.g. MC/AGQ [7]) for discrete response models and robust design optimisation accounting for model and/or parameter uncertainty [8,9]
- To submit PFIM package to CRAN
- To increase the interoperability of PFIM with estimation parameter softwares, as aimed by the DDMoRe project.

References
[1] Durrant, C., Lantier, C., le Nagard, H., Menétré, F., Cochet, F., Nguyen, T.T. et al. PFIM 4.0: an extended R program for design evaluation and optimization for nonlinear mixed-effects models. *Comput Methods Programs Biomed*. 2018;156:217-29. [2] Chastelain, JM. Object-Oriented Programming. *Practical Programming with R*. San Jose, CA: No Starch Press; 2013. [3] Chastelain, JM. Object-Oriented Programming. *Practical Programming with R*. San Jose, CA: No Starch Press; 2013. [4] Chastelain, JM, Nguyen, T.T, Tang, Y, Menétré, F. Population-based optimal design using the Fisher information matrix for nonlinear mixed-effects models with evaluation of prior distributions. *Pharm Res*. 2015;32:1555-67. [5] Tang, Y, Cochet, F, Nguyen, T.T, Menétré, F. Bayesian shrinkage prediction for nonlinear mixed-effects models using Monte Carlo sampling and adaptive Gaussian quadrature. *Comput Stat Data Anal*. 2017;61:1205-19. [6] Tang, Y, Cochet, F, Nguyen, T.T, Menétré, F. Robust design in longitudinal studies using the Fisher information matrix for nonlinear mixed-effects models with varying covariates. *Stat Methods Med Res*. 2019;29:111-17. [7] Tang, Y, Cochet, F, Nguyen, T.T, Menétré, F. Robust design in longitudinal studies using the Fisher information matrix for nonlinear mixed-effects models with varying covariates. *Stat Methods Med Res*. 2019;29:111-17. [8] Tang, Y, Cochet, F, Nguyen, T.T, Menétré, F. Robust design in longitudinal studies using the Fisher information matrix for nonlinear mixed-effects models with varying covariates. *Stat Methods Med Res*. 2019;29:111-17. [9] Tang, Y, Cochet, F, Nguyen, T.T, Menétré, F. Robust design in longitudinal studies using the Fisher information matrix for nonlinear mixed-effects models with varying covariates. *Stat Methods Med Res*. 2019;29:111-17.

Titre : Modélisation de données longitudinales : protocoles robustes dans les essais cliniques, interactions bactériophages/bactéries

Résumé : En amont de la modélisation de données d'études longitudinales par modèles non linéaires à effets mixtes (MNLEM), il est nécessaire de planifier des protocoles informatifs. Une approche efficace se base sur l'utilisation de la matrice d'information de Fisher (FIM). L'évaluation et l'optimisation de protocoles nécessite des connaissances *a priori* sur le modèle. Si le modèle est finalement différent du modèle supposé *a priori*, cela peut conduire à un protocole sous-optimal. Le premier travail de thèse propose une approche de protocole robuste tenant compte de l'incertitude du modèle. Différents algorithmes sont utilisés pour l'optimisation de protocole basée sur la FIM. Les algorithmes multiplicatifs permettent l'optimisation de protocoles en s'appuyant sur une discrétisation de l'espace et donner un poids plus important aux protocoles les plus informatifs. Ce type d'algorithme a démontré son efficacité pour des modèles de régression individuelle. Dans le cadre de cette thèse, ils ont été appliqués aux MNLEM. L'utilisation de simulations d'essais cliniques (CTS) permet de comparer un nombre restreint de protocoles. Un projet d'application a été conduit en immuno-oncologie pour l'étude d'une combinaison de deux médicaments en phase précoce de développement clinique. Le critère de jugement utilisé était basé sur les tailles de tumeurs individuelles prédites à partir d'un modèle d'inhibition de la croissance tumorale. Dans ce cadre, des protocoles à un, deux ou quatre bras de traitement ont été comparés par CTS. Dans le contexte de la propagation des bactéries multi-résistantes, les bactériophages (des virus qui n'infectent que les bactéries) suscitent de plus en plus d'intérêt. Il est néanmoins nécessaire de déterminer les modalités optimales d'administration des phages avant que l'on puisse proposer un traitement chez l'homme. Une série d'expériences *in vitro* et *in vivo* ont été conduites et l'ensemble des données a permis de construire un modèle mathématique qui caractérise l'interaction phage-bactérie et qui a permis de quantifier l'efficacité de la phagothérapie en fonction de la dose et de la voie d'administration.

Mots clefs : modèles non linéaires à effets mixtes ; études longitudinales ; optimisation de protocoles ; matrice d'Information de Fisher ; algorithmes multiplicatifs ; modèles d'inhibition de la croissance tumorale ; bactériophages ; modèles de cinétique bactérienne ; modèles de dynamique virale

Title: Longitudinal data modeling: robust protocols in clinical trials, bacteriophages/bacteria interactions

Abstract: Prior to data modeling of longitudinal studies using nonlinear mixed effects models (NLMEM), informative protocols need to be planned. An effective approach is based on the use of the Fisher Information Matrix (FIM). The evaluation and optimization of protocols requires a priori knowledge of the model. If the model is ultimately different from the *a priori* assumed model, this may lead to a sub-optimal protocol. The first thesis work proposes a robust design approach taking into account the model uncertainty. Different algorithms are used for design optimization based on FIM. Multiplicative algorithms allow design optimization based on discretization of the design space and give more importance to the most informative protocols. This type of algorithm has demonstrated its effectiveness for individual regression models. In this thesis, they have been applied to NLMEM. The use of clinical trial simulations (CTS) allows the comparison of a small number of designs. An application project was conducted in immuno-oncology for the study of a combination of two drugs in early clinical development. The endpoint used was based on individual tumor sizes predicted from a tumor growth inhibition model. Designs with one, two or four treatment arms were compared by CTS. In the context of the spread of multi-resistant bacteria, bacteriophages (viruses that only infect bacteria) are of increasing interest. Nevertheless, it is necessary to determine the optimal phage administration scheme before treatment can be proposed in humans. A series of *in vitro* and *in vivo* experiments have been conducted and all the data has made it possible to build a mathematical model that characterizes the phage-bacteria interaction and that quantifies the effectiveness of phagotherapy depending on the dose and the route of administration.

Keywords: nonlinear mixed effects models ; longitudinal studies ; optimal design ; Fisher information matrix ; multiplicative algorithms ; tumor growth inhibition models ; bacteriophages ; bacterial kinetics models ; viral dynamics models