

2024-2025

Thèse

pour le

Diplôme d'État de Docteur en Pharmacie

**Développement d'une stratégie
d'échantillonnage limité pour des
nanocapsules lipidiques (NCL)**

--

**Development of a limited sampling strategy
for lipid nanocapsules (LNCs)**

CISSÉ Jean-Luc

Né le 15 mars 2001 à Angers

Sous la direction de **LEBRETON Vincent**

Membres du jury

| Saulnier Patrick (Président)

| Lebreton Vincent (Directeur)

| Legeay Samuel (Co-Directeur)

| Chatelut Étienne (Membre)

| Montarnal Aurélie (Membre)



**FACULTÉ
DE SANTÉ**

UNIVERSITÉ D'ANGERS

Soutenue publiquement le :
17 juin 2025 à Angers

ENGAGEMENT DE NON PLAGIAT

Je, soussigné Cissé Jean-Luc déclare être pleinement conscient que le plagiat de documents ou d'une partie d'un document publiée sur toutes formes de support, y compris l'internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.

En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce rapport ou mémoire.

signé par l'étudiant(e) le **17 / 11 / 2023**

LISTE DES ENSEIGNANTS DE LA FACULTÉ DE SANTÉ D'ANGERS

Doyen de la Faculté : Pr Cédric ANNWEILER

Vice-Doyen de la Faculté et directeur du département de pharmacie : Pr Sébastien FAURE

Directeur du département de médecine : Pr Vincent DUBEE

PROFESSEURS DES UNIVERSITÉS

ABRAHAM Pierre	PHYSIOLOGIE	Médecine
ANGOULVANT Cécile	MEDECINE GENERALE	Médecine
ANNWEILER Cédric	GERIATRIE ET BIOLOGIE DU VIEILLISSEMENT	Médecine
ASFAR Pierre	REANIMATION	Médecine
AUBE Christophe	RADIOLOGIE ET IMAGERIE MEDICALE	Médecine
AUGUSTO Jean-François	NEPHROLOGIE	Médecine
BAUFRETON Christophe	CHIRURGIE THORACIQUE ET CARDIOVASCULAIRE	Médecine
BELLANGER William	MEDECINE GENERALE	Médecine
BELONCLE François	REANIMATION	Médecine
BIERE Loïc	CARDIOLOGIE	Médecine
BIGOT Pierre	UROLOGIE	Médecine
BONNEAU Dominique	GENETIQUE	Médecine
BOUCHARA Jean-Philippe	PARASITOLOGIE ET MYCOLOGIE	Médecine
BOUET Pierre-Emmanuel	GYNECOLOGIE-OBSTETRIQUE	Médecine
BOURSIER Jérôme	GASTROENTEROLOGIE ; HEPATOLOGIE	Médecine
BOUVARD Béatrice	RHUMATOLOGIE	Médecine
BRIET Marie	PHARMACOLOGIE	Médecine
CAMPONE Mario	CANCEROLOGIE ; RADIOTHERAPIE	Médecine
CAROLI-BOSC François-Xavier	GASTROENTEROLOGIE ; HEPATOLOGIE	Médecine
CASSEREAU Julien	NEUROLOGIE	Médecine
CLERE Nicolas	PHARMACOLOGIE / PHYSIOLOGIE	Pharmacie
COLIN Estelle	GENETIQUE	Médecine
CONNAN Laurent	MEDECINE GENERALE	Médecine
COPIN Marie-Christine	ANATOMIE ET CYTOLOGIE PATHOLOGIQUES	Médecine
COUTANT Régis	PEDIATRIE	Médecine
CUSTAUD Marc-Antoine	PHYSIOLOGIE	Médecine
CRAUSTE-MANCIET Sylvie	PHARMACOTECHNIE HOSPITALIERE	Pharmacie
DE CASABIANCA Catherine	MEDECINE GENERALE	Médecine
DERBRE Séverine	PHARMACOGNOSIE	Pharmacie
DESCAMPS Philippe	GYNECOLOGIE-OBSTETRIQUE	Médecine
D'ESCATHA Alexis	MEDECINE ET SANTE AU TRAVAIL	Médecine
DINOMAIS Mickaël	MEDECINE PHYSIQUE ET DE READAPTATION	Médecine
DUBEE Vincent	MALADIES INFECTIEUSES ET TROPICALES	Médecine
DUCANCELLE Alexandra	BACTERIOLOGIE-VIROLOGIE ; HYGIENE HOSPITALIERE	Médecine
DUVERGER Philippe	PEDOPSYCHIATRIE	Médecine
EVEILLARD Matthieu	BACTERIOLOGIE-VIROLOGIE	Pharmacie
FAURE Sébastien	PHARMACOLOGIE PHYSIOLOGIE	Pharmacie
FOURNIER Henri-Dominique	ANATOMIE	Médecine
FOUQUET Olivier	CHIRURGIE THORACIQUE ET CARDIOVASCULAIRE	Médecine
FURBER Alain	CARDIOLOGIE	Médecine
GAGNADOUX Frédéric	PNEUMOLOGIE	Médecine
GOHIER Bénédicte	PSYCHIATRIE D'ADULTES	Médecine
GUARDIOLA Philippe	HEMATOLOGIE ; TRANSFUSION	Médecine
GUILLET David	CHIMIE ANALYTIQUE	Pharmacie
HUNAUULT-BERGER Mathilde	HEMATOLOGIE ; TRANSFUSION	Médecine



FACULTÉ DE SANTÉ

UNIVERSITÉ D'ANGERS

JEANNIN Pascale	IMMUNOLOGIE	Médecine
KAZOUR François	PSYCHIATRIE	Médecine
KEMPF Marie	BACTERIOLOGIE-VIROLOGIE ; HYGIENE HOSPITALIERE	Médecine
KUN-DARBOIS Daniel	CHIRURGIE MAXILLO-FACIALE ET STOMATOLOGIE	Médecine
LACOEUILLE FRANCK	RADIOPHARMACIE	Pharmacie
LACOURREYE Laurent	OTO-RHINO-LARYNGOLOGIE	Médecine
LAGARCE Frédéric	BIOPHARMACIE	Pharmacie
LANDREAU Anne	BOTANIQUE/ MYCOLOGIE	Pharmacie
LASOCKI Sigismond	ANESTHESIOLOGIE-REANIMATION	Médecine
LEBDAI Souhil	UROLOGIE	Médecine
LEGENDRE Guillaume	GYNECOLOGIE-OBSTETRIQUE	Médecine
LEGRAND Erick	RHUMATOLOGIE	Médecine
LEMEE Jean-Michel	NEUROCHIRURGIE	Médecine
LERMITE Emilie	CHIRURGIE GENERALE	Médecine
LEROLLE Nicolas	REANIMATION	Médecine
LIBOUBAN Hélène	HISTOLOGIE	Médecine
LUQUE PAZ Damien	HEMATOLOGIE BIOLOGIQUE	Médecine
MARCHAIS Véronique	BACTERIOLOGIE-VIROLOGIE	Pharmacie
MARTIN Ludovic	DERMATO-VERNEROLOGIE	Médecine
MAY-PANLOUP Pascale	BIOLOGIE ET MEDECINE DU DEVELOPPEMENT ET DE LA REPRODUCTION	Médecine
MENEI Philippe	NEUROCHIRURGIE	Médecine
MERCAT Alain	REANIMATION	Médecine
ORVAIN Corentin	HEMATOLOGIE ; TRANSFUSION	Médecine
PAISANT Anita	RADIOLOGIE	Médecine
PAPON Nicolas	PARASITOLOGIE ET MYCOLOGIE MEDICALE	Pharmacie
PASSIRANI Catherine	CHIMIE GENERALE	Pharmacie
PELLIER Isabelle	PEDIATRIE	Médecine
PETIT Audrey	MEDECINE ET SANTE AU TRAVAIL	Médecine
PICQUET Jean	CHIRURGIE VASCULAIRE ; MEDECINE VASCULAIRE	Médecine
PODEVIN Guillaume	CHIRURGIE INFANTILE	Médecine
PROCACCIO Vincent	GENETIQUE	Médecine
PRUNIER Delphine	BIOCHIMIE ET BIOLOGIE MOLECULAIRE	Médecine
PRUNIER Fabrice	CARDIOLOGIE	Médecine
PY Thibaut	MEDECINE GENERALE	Médecine
RAMOND-ROQUIN Aline	MEDECINE GENERALE	Médecine
REYNIER Pascal	BIOCHIMIE ET BIOLOGIE MOLECULAIRE	Médecine
RIOU Jérémie	BIOSTATISTIQUE	Pharmacie
RINEAU Emmanuel	ANESTHESIOLOGIE REANIMATION	Médecine
RIQUIN Elise	PEDOPSYCHIATRIE ; ADDICTOLOGIE	Médecine
RODIEN Patrice	ENDOCRINOLOGIE, DIABETE ET MALADIES METABOLIQUES	Médecine
ROQUELAURE Yves	MEDECINE ET SANTE AU TRAVAIL	Médecine
ROUGE-MAILLART Clotilde	MEDECINE LEGALE ET DROIT DE LA SANTE	Médecine
ROUSSEAU Audrey	ANATOMIE ET CYTOLOGIE PATHOLOGIQUES	Médecine
ROUSSEAU Pascal	CHIRURGIE PLASTIQUE, RECONSTRUCTRICE ET ESTHETIQUE	Médecine
ROUSSELET Marie-Christine	ANATOMIE ET CYTOLOGIE PATHOLOGIQUES	Médecine
ROY Pierre-Marie	MEDECINE D'URGENCE	Médecine
SAULNIER Patrick	BIOPHYSIQUE ET BIOSTATISTIQUES	Pharmacie
SERAPHIN Denis	CHIMIE ORGANIQUE	Pharmacie
SCHMIDT Aline	HEMATOLOGIE ; TRANSFUSION	Médecine
TESSIER-CAZENEUVE Christine	MEDECINE GENERALE	Médecine
TRZEPIZUR Wojciech	PNEUMOLOGIE	Médecine
UGO Valérie	HEMATOLOGIE ; TRANSFUSION	Médecine



FACULTÉ DE SANTÉ

UNIVERSITÉ D'ANGERS

URBAN Thierry	PNEUMOLOGIE	Médecine
VAN BOGAERT Patrick	PEDIATRIE	Médecine
VENARA Aurélien	CHIRURGIE VISCERALE ET DIGESTIVE	Médecine
VENIER-JULIENNE Marie-Claire	PHARMACOTECHNIE	Pharmacie
VERNY Christophe	NEUROLOGIE	Médecine
WILLOTEAUX Serge	RADIOLOGIE ET IMAGERIE MEDICALE	Médecine

MAÎTRES DE CONFÉRENCES

AMMI Myriam	CHIRURGIE VASCULAIRE ET THORACIQUE	Médecine
BAGLIN Isabelle	CHIMIE THERAPEUTIQUE	Pharmacie
BASTIAT Guillaume	BIOPHYSIQUE ET BIOSTATISTIQUES	Pharmacie
BEAUVILLAIN Céline	IMMUNOLOGIE	Médecine
BEGUE Cyril	MEDECINE GENERALE	Médecine
BELIZNA Cristina	MEDECINE INTERNE	Médecine
BERNARD Florian	ANATOMIE	Médecine
BESSAGUET Flavien	PHYSIOLOGIE PHARMACOLOGIE	Pharmacie
BLANCHET Odile	HEMATOLOGIE ; TRANSFUSION	Médecine
BOISARD Séverine	CHIMIE ANALYTIQUE	Pharmacie
BOUCHER Sophie	ORL	Médecine
BRIET Claire	ENDOCRINOLOGIE, DIABETE ET MALADIES METABOLIQUES	Médecine
BRILLAND Benoit	NEPHROLOGIE	Médecine
BRIS Céline	BIOCHIMIE ET BIOLOGIE MOLECULAIRE	Pharmacie
BRUGUIERE Antoine	PHARMACOGNOSIE	Pharmacie
CAPITAIN Olivier	CANCEROLOGIE ; RADIOTHERAPIE	Médecine
CHABRUN Floris	BIOCHIMIE ET BIOLOGIE MOLECULAIRE	Pharmacie
CHAO DE LA BARCA Juan-Manuel	BIOCHIMIE ET BIOLOGIE MOLECULAIRE	Médecine
CHOPIN Matthieu	MEDECINE GENERALE	
CODRON Philippe	NEUROLOGIE	Médecine
DEMAS Josselin	SCIENCES DE LA READAPTATION	Médecine
DESHAYES Caroline	BACTERIOLOGIE VIROLOGIE	Pharmacie
DOUILLET Delphine	MEDECINE D'URGENCE	Médecine
FERRE Marc	BIOLOGIE MOLECULAIRE	Médecine
FORTRAT Jacques-Olivier	PHYSIOLOGIE	Médecine
GHALI Maria	MEDECINE GENERALE	Médecine
GUELFJ Jessica	MEDECINE GENERALE	Médecine
HADJ MAHMOUD Dorra	IMMUNOLOGIE	Pharma
HAMEL Jean-François	BIOSTATISTIQUES, INFORMATIQUE MEDICALE	Médicale
HAMON Cédric	MEDECINE GENERALE	Médecine
HELESBEUX Jean-Jacques	CHIMIE ORGANIQUE	Pharmacie
HERIVAUX Anaïs	BIOTECHNOLOGIE	Pharmacie
HINDRE François	BIOPHYSIQUE	Médecine
JOUSSET-THULLIER Nathalie	MEDECINE LEGALE ET DROIT DE LA SANTE	Médecine
JUDALET-ILLAND Ghislaine	MEDECINE GENERALE	Médecine
KHIATI Salim	BIOCHIMIE ET BIOLOGIE MOLECULAIRE	Médecine
LEFEUVRE Caroline	BACTERIOLOGIE ; VIROLOGIE	Médecine
LEGEAY Samuel	PHARMACOCINETIQUE	Pharmacie
LEPELTIER Elise	CHIMIE GENERALE	Pharmacie
LETOURNEL Franck	BIOLOGIE CELLULAIRE	Médecine
MABILLEAU Guillaume	HISTOLOGIE, EMBRYOLOGIE ET CYTOGENETIQUE	Médecine
MALLET Sabine	CHIMIE ANALYTIQUE	Pharmacie
MAROT Agnès	PARASITOLOGIE ET MYCOLOGIE MEDICALE	Pharmacie
MESLIER Nicole	PHYSIOLOGIE	Médecine
MIOT Charline	IMMUNOLOGIE	Médecine
MOUILLIE Jean-Marc	PHILOSOPHIE	Médecine



FACULTÉ DE SANTÉ

UNIVERSITÉ D'ANGERS

NAIL BILLAUD Sandrine	IMMUNOLOGIE	Pharmacie
PAILHORIES Hélène	BACTERIOLOGIE-VIROLOGIE	Médecine
PAPON Xavier	ANATOMIE	Médecine
PASCO-PAPON Anne	RADIOLOGIE ET IMAGERIE MEDICALE	Médecine
PENCHAUD Anne-Laurence	SOCIOLOGIE	Médecine
PIHET Marc	PARASITOLOGIE ET MYCOLOGIE	Médecine
PIRAUX Arthur	OFFICINE	Pharmacie
POIROUX Laurent	SCIENCES INFIRMIERES	Médecine
RONY Louis	CHIRURGIE ORTHOPEDIQUE ET TRAUMATOLOGIQUE	Médecine
ROGER Emilie	PHARMACOTECHNIE	Pharmacie
SAVARY Camille	PHARMACOLOGIE-TOXICOLOGIE	Pharmacie
SCHMITT Françoise	CHIRURGIE INFANTILE	Médecine
SCHINKOWITZ Andréas	PHARMACOGNOSIE	Pharmacie
SPIESSER-ROBELET Laurence	PHARMACIE CLINIQUE ET EDUCATION THERAPEUTIQUE	Pharmacie
TEXIER-LEGENDRE Gaëlle	MEDECINE GENERALE	Médecine
VIAULT Guillaume	CHIMIE ORGANIQUE	Pharmacie

AUTRES ENSEIGNANTS

ATER		
BARAKAT Fatima	CHIMIE ANALYTIQUE	Pharmacie
ATCHADE Constantin	GALENIQUE	Pharmacie
PRCE		
AUTRET Erwan	ANGLAIS	Santé
BARBEROUSSE Michel	INFORMATIQUE	Santé
COYNE Ashley	ANGLAIS	Santé
O'SULLIVAN Kayleigh	ANGLAIS	Santé
RIVEAU Hélène	ANGLAIS	Santé
PAST-MAST		
AUBRUCHET Hélène	OFFICINE	Pharmacie
BEAUVAIS Vincent	OFFICINE	Pharmacie
BRAUD Cathie	OFFICINE	Pharmacie
CAVAILLON Pascal	PHARMACIE INDUSTRIELLE	Pharmacie
CHAMPAGNE Romain	MEECINE PHYSIQUE ET READAPTATION	Médecine
DILÉ Nathalie	OFFICINE	Pharmacie
GUILLET Anne-Françoise	PHARMACIE DEUST PREPARATEUR	Pharmacie
GUITTON Christophe	MEDECINE INTENSIVE-REANIMATION	Médecine
KAASSIS Mehdi	GASTRO-ENTEROLOGIE	Médecine
LAVIGNE Christian	MEDECINE INTERNE	Médecine
MARSAN-POIROUX	COMMUNICATION	Pharmacie
MOAL Frédéric	PHARMACIE CLINIQUE	Pharmacie
PICCOLI Giorgina	NEPHROLOGIE	Médecine
POMMIER Pascal	CANCEROLOGIE-RADIOTHERAPIE	Médecine
SAVARY Dominique	MEDECINE D'URGENCE	Médecine
PLP		
CHIKH Yamina	ECONOMIE-GESTION	Médecine
AHU		
CORVAISIER Mathieu	PHARMACIE CLINIQUE	Pharmacie
ROBIN Julien	DISPOSITIFS MEDICAUX	Pharmacie

REMERCIEMENTS

À mon directeur de thèse Monsieur Lebreton Vincent, un grand merci de m'avoir proposé ce sujet et surtout de m'avoir fait confiance. Merci pour votre accompagnement tout au long de la thèse ainsi que pour votre ouverture d'esprit vis-à-vis des différentes idées évoquées. Ce fût un travail long et difficile, je suis très heureux d'avoir collaboré avec vous.

À mon co-directeur de thèse Monsieur Legeay Samuel, merci de tout cœur de m'avoir accompagné durant mon stage. Depuis mes années à la faculté de Pharmacie d'Angers, j'ai toujours été impressionné par votre façon d'enseigner ainsi que par votre pédagogie. Merci pour votre temps accordé. Je suis très heureux d'avoir travaillé avec vous, un grand merci.

À Monsieur Saulnier Patrick, en tant que président de thèse. Je vous remercie d'avoir accepté la présidence de cette soutenance, merci pour le temps que vous y consacrez.

Aux membres du jury, je vous adresse mes plus sincères remerciements pour avoir accepté d'être membre du jury d'évaluation de ma thèse. C'est un grand honneur pour moi de présenter mes travaux devant vous.

À mes professeurs de l'École de l'IMT Mines Albi, Monsieur Dirion, Madame Montarnal, je tenais également à vous remercier pour ces deux années passées avec vous à l'École des Mines d'Albi. Merci pour vos enseignements de très bonne qualité ainsi que pour les notions enseignées. C'est en grande partie grâce à vos enseignements que j'ai pu réaliser ce travail, merci de tout cœur.

À ma famille, merci de m'avoir soutenu pendant ces longues années d'études. Merci maman de m'avoir donné la motivation et la détermination dans mon parcours. Je t'aime très fort et je suis très fier de présenter ce travail devant toi, je suis très fier d'avoir une maman en or comme toi. Merci maman.

Papa, quand j'étais petit je t'ai dit un jour que je travaillerais dur pour réussir mes études, pour que tu sois fier de moi. Ce jour-là, on a commencé à travailler ensemble. Puis tu m'as offert le livre du corps humain ce qui a créé en moi l'amour pour la médecine. Ensuite, tu as décidé de me transmettre toutes tes connaissances et tu m'as donné l'amour des sciences. Aujourd'hui tu n'es plus là. Mais j'ai continué à me battre et j'ai continué notre travail. Aujourd'hui, avec cette thèse j'ai atteint notre objectif. Et c'est grâce à toi papa... Tu me manques profondément, je t'aime très très fort.

Liste des abréviations.....	11
Introduction	12
1. Nanomédicaments et nanocapsules	13
1.1. Définition	13
1.2. Classification des nanomédicaments	14
1.2.1. Les formes encapsulées	15
1.2.1.1. Dendrimères	15
1.2.1.2. Micelles	16
1.2.1.3. Liposomes	16
1.2.1.4. Nanocapsules	17
1.2.2. Les formes conjuguées	17
1.2.2.1. Les anticorps conjugués aux principes actifs (<i>antibody drug conjugates</i>)	17
1.2.2.2. Médicaments liés à des nanocapsules d'albumine (<i>nab-drugs</i>)	18
1.3. Applications médicales	18
1.3.1. Cancérologie	18
1.3.2. Cardiologie	19
1.3.3. Infectiologie	19
1.3.4. Neurologie	19
1.3.5. Maladies auto-immunes	19
2. La pharmacocinétique des nanomédicaments	20
2.1. Propriétés pharmacocinétiques des nanomédicaments	20
2.1.1. Absorption	20
2.1.2. Distribution	21
2.1.2.1. L'effet EPR	22
2.1.2.2. Ciblage actif	24
2.1.2.3. Barrière à la distribution des nanomédicaments	24
2.1.3. Métabolisme	24
2.1.4. Élimination	25
2.1.4.1. L'instabilité dans les fluides biologiques	25
2.1.4.2. L'élimination hépatobiliaire	26
2.1.4.3. Élimination rénale	27
2.1.4.4. L'élimination par le système lymphatique	28
2.1.5. Influence des propriétés physico-chimiques sur leur pharmacocinétique des nanomédicaments	28
2.1.5.1. Taille	28
2.1.5.2. Composition	30
2.1.5.3. Charge électrique	31
2.1.5.4. Forme	31
2.1.6. Influence des variabilités interindividuelles sur la pharmacocinétique des nanomédicaments	32
2.1.6.1. L'âge	32
2.1.6.2. Le poids	33
2.1.6.3. Variabilité physio-pathologique	33
2.1.6.4. Importance du système immunitaire	33
2.2. Paramètres et modèles pharmacocinétiques	34
2.2.1. Paramètres pharmacocinétiques	34
2.2.1.1. Paramètres graphiques	35
2.2.1.1.1. Aire sous la courbe (AUC)	35
2.2.1.1.2. Le temps de demi-vie d'élimination	37
2.2.1.1.3. Le temps de résidence moyen (MRT)	37
2.2.1.2. Paramètres mécanistiques	38
2.2.1.2.1. La clairance d'élimination totale	38
2.2.1.2.1.1. Clairance rénale	39
2.2.1.2.1.2. Clairance métabolique hépatique	40
2.2.1.2.2. Le volume de distribution	40
2.2.2. Modèle pharmacocinétique non linéaire et modèle populationnel	42
2.2.2.1. Modèle non linéaire	42
2.2.2.1.1. Métabolisme saturable	42

2.2.2.1.2.	Sécrétion tubulaire rénale	43
2.2.2.1.3.	Fixation aux protéines plasmatiques	43
2.2.2.2.	Modèle populationnel.....	43
2.3.	Intérêt de la technique FRET pour l'évaluation pharmacocinétique des NCL.....	44
2.3.1.	Analyse non compartimentale.....	47
2.3.2.	Modèle de population	48
3.	Limited Sampling Strategies (LSS).....	49
3.1.	Principe général.....	49
3.2.	Traitement des données.....	50
3.2.1.	Exploration des données	50
3.2.1.1.	Statistiques exploratoires.....	50
3.2.1.1.1.	Effectif	50
3.2.1.1.2.	Moyenne arithmétique	51
3.2.1.1.3.	Médiane	52
3.2.1.1.4.	Valeur minimale et maximale	53
3.2.1.1.5.	Variance.....	53
3.2.1.1.6.	Écart type	53
3.2.1.1.7.	Quartiles	54
3.2.1.2.	Visualisation des données	55
3.2.1.2.1.	Nuage de points	55
3.2.1.2.2.	Histogrammes	55
3.2.1.2.3.	Box plots.....	56
3.2.2.	Traitement des valeurs manquantes	57
3.2.2.1.	Imputation par une constante	57
3.2.2.2.	Imputation par la moyenne des valeurs.....	59
3.2.2.3.	Imputation par la médiane des valeurs.....	59
3.2.3.	Mise à l'échelle	60
3.2.3.1.	Standardisation	60
3.2.3.2.	Normalisation.....	62
3.2.4.	Transformation de variables	62
3.2.4.1.	Transformation Box-Cox.....	63
3.2.4.2.	Transformation puissance	63
3.2.4.3.	Transformation logarithmique	63
3.2.5.	Analyse en composante principale	63
3.2.6.	Sélection de variables	64
3.2.6.1.	Coefficient de corrélation.....	64
3.2.6.1.1.	Coefficient de Bravais-Pearson	65
3.2.6.1.2.	Coefficient de Spearman	65
3.2.6.2.	Méthode exhaustive	66
3.2.6.3.	Méthodes pas à pas	66
3.2.6.3.1.	Sélection ascendante	66
3.2.6.3.2.	Sélection descendante	67
3.2.6.3.3.	Sélection <i>stepwise</i>	68
3.3.	Les modèles	68
3.3.1.	La régression linéaire multiple.....	68
3.3.1.1.	Principe.....	68
3.3.1.2.	Modélisation	69
3.3.2.	La modélisation bayésienne	70
3.3.2.1.	Principe.....	70
3.3.2.2.	Modélisation	71
3.3.2.2.1.	Modèle de structure	71
3.3.2.2.1.1.	Modèle mono-compartmental.....	72
3.3.2.2.1.2.	Modèle bi-compartmental	73
3.3.2.2.1.3.	Modèle tri-compartmental.....	75
3.3.2.2.2.	Modèles statistiques.....	75
3.3.2.2.2.1.	Modèle décrivant la variabilité interindividuelle	75
3.3.2.2.2.2.	Modèle décrivant la variabilité résiduelle	76
3.3.2.2.3.	Recherche de covariables.....	77
3.3.2.2.4.	Estimation des paramètres	77
3.3.3.	Comparaison entre la régression linéaire multiple et la modélisation bayésienne.....	79

3.4.	La validation des modèles	79
3.4.1.	La validation de la régression linéaire multiple	79
3.4.1.1.	Validation par les métriques	79
3.4.1.1.1.	Le coefficient de détermination (r^2)	79
3.4.1.1.2.	Le coefficient de détermination ajusté ($r^2_{\text{Ajusté}}$)	80
3.4.1.1.3.	La <i>Root Mean Squared Error</i>	80
3.4.1.1.4.	La <i>Mean Predictive Error</i>	80
3.4.1.1.5.	La métrique J	81
3.4.1.2.	Validation croisée	82
3.4.1.2.1.	Cross validation <i>K-folds</i>	82
3.4.1.2.2.	<i>Leave one out cross validation</i> (LOOCV)	83
3.4.1.3.	Validation des hypothèses	84
3.4.1.3.1.	Graphique des valeurs prédites en fonction des valeurs observées (<i>predicted vs observed</i>)	84
3.4.1.3.2.	Linéarité du modèle	85
3.4.1.3.3.	Homoscédasticité des résidus	85
3.4.1.3.4.	Normalité des résidus	86
3.4.1.3.5.	Non-autocorrélation des résidus	86
3.4.1.3.6.	Non multicolinéarité entre les variables	86
3.4.1.3.7.	Bland-Altman	87
3.4.1.3.8.	Significativité des variables et du modèle	87
3.4.2.	La validation de la modélisation bayésienne	87
3.4.2.1.	Validation par les métriques	87
3.4.2.2.	Validation par le <i>bootstrap</i>	88
3.4.2.3.	Validation graphique	88
3.5.	Quelques exemples d'application médicales	89
4.	La LSS appliquée aux LNCs	90
4.1.	Objectifs	90
4.2.	Matériels et Méthodes	91
4.3.	Résultats	95
4.3.1.	Analyse descriptive des données	95
4.3.1.1.	Analyse exploratoire	95
4.3.1.2.	Paramètres pharmacocinétiques	96
4.3.1.3.	Boîtes à moustaches	96
4.3.1.4.	Coefficient de corrélation de Bravais-Pearson et de Spearman	97
4.3.2.	Résultats LSS à 1 point	98
4.3.3.	Résultats LSS à 2 points	98
4.3.4.	Résultats LSS à 3 points	99
4.4.	Validation du modèle choisi	100
4.4.1.	Coefficients du modèle et composition des composantes principales sélectionnées	100
4.4.2.	<i>Predicted vs observed</i>	101
4.4.3.	Vérification de la linéarité du modèle	101
4.4.4.	Vérification de la normalité des résidus	101
4.4.5.	Vérification de l'homoscédasticité des résidus	102
4.4.6.	Vérification de la non-autocorrélation des résidus	102
4.4.7.	Vérification de la non-multicolinéarité	102
4.4.8.	Bland-Altman	103
4.4.9.	Significativité du modèle	104
4.5.	Discussion	104
	Conclusion	107
	Bibliographie	108
	Annexes	113
	Table des illustrations	167
	Table des tableaux	169
	Serment de Galien	170
	Résumé	171

Liste des abréviations

ACP	Analyse en composantes principales
ADC	<i>Antibody Drug</i>
ADME	Absorption Distribution Métabolisme Élimination
AUC	<i>Area Under the Curve</i>
AUMC	<i>Area Under the Moment Curve</i>
BLOQ	<i>Below the Lower limit Of Quantification</i>
CL	Clairance
DiD	1,1' -dioctadecyl-3,3,3' ,3' -tetramethylindo-dicarboyanine perchlorate
Dil	1,10-dioctadecyl-3,3,30,30-tetramethyl-indocarboyanine perchlorate
EIQ	Étendu Interquartile
FDA	<i>Food and Drug Administration</i>
FRET	<i>Förster resonance energy transfer</i>
IPRED	<i>Individual predicted concentrations</i>
IV	Intraveineux
K _d	Constante de dissociation
kDa	Kilo Dalton
k _e	Constante d'élimination
LLOQ	<i>Lower Limit Of Quantification</i>
NCL	Nanocapsules lipidiques
LOOCV	<i>Leave-one-out Cross Validation</i>
LSS	<i>Limited Sampling Strategies</i>
MPE	<i>Mean Predictive Error</i>
MRT	<i>Mean retention time</i>
Norm	Normale
OFV	<i>Objective Function Value</i>
PC	<i>Principal Component</i>
PEG	Pégylé
PHDCA	Polyhexadecylcyanoacrylate
pH	Potentiel Hydrogène
PK	Pharmacocinétique
PR	<i>Proximity Ratio</i>
RFECV	<i>Recursive Feature Elimination Cross Validation</i>
RMSE	<i>Root Mean Squared Error</i>
RSE	<i>Relative Standard Error</i>
SA	Stéarylaminé
SPM	Système phagocytaire mononucléé
V _d	Volume de distribution
VIF	<i>Variance Inflation Factor</i>
VPC	<i>Visual Predictive Check</i>

Introduction

Les nanomédicaments sont des structures de faible taille de l'ordre de quelques dizaines de nanomètres constitués de principe(s) actif(s) encapsulé(s) dans une nanocapsule. Ils ont la propriété de pouvoir vectoriser le ou les principe(s) actif(s) directement vers leur(s) cible(s) ceci permettant alors d'augmenter leur biodisponibilité et de limiter les effets indésirables.

Il est possible de contrôler un certain nombre de paramètres liés à la formulation de ces nanomédicaments tels que la taille, la forme ou encore la composition de ces derniers. Ces variations peuvent avoir pour conséquence d'influencer leur activité pharmacologique, leur passage à travers les membranes cellulaires ou encore leur détection et élimination par le système phagocytaire mononucléé (SPM).

Les nanomédicaments représentent donc une avancée importante pour l'amélioration des traitements actuels. Ils peuvent être utilisés dans de nombreuses applications thérapeutiques. C'est notamment le cas en oncologie que les nanomédicaments semblent particulièrement intéressants pour améliorer le ciblage des cellules tumorales. Ce ciblage, permet de réduire les nombreux effets indésirables causés par un manque de sélectivité des traitements traditionnels vis-à-vis des cellules tumorales.

Cependant, le ciblage spécifique des nanomédicaments est théorique car en pratique seulement 0,3 % à 3% de la quantité de nanomédicaments administrée atteint son site d'action¹. Ce constat apparaît comme l'une des limites majeures à l'essor des nanomédicaments en clinique. Pour l'expliquer, il semble important de se pencher sur la pharmacocinétique des nanomédicaments.

La pharmacocinétique des nanomédicaments est particulière et reste encore peu étudiée à ce jour. L'un des points bloquants des études pharmacocinétiques à l'heure actuelle, réside dans la nécessité d'effectuer de nombreux prélèvements sanguins qui s'avèrent chronophages, coûteux et douloureux. Ces inconvénients sont autant de freins au développement mais pourraient être levés si le nombre de point à prélever était réduit.

Pour réduire le nombre de prélèvement, une technique statistique nommée *Limited Sampling Strategies* (LSS) peut être utilisée. Cette technique identifie les temps de prélèvements optimaux afin d'estimer avec précision et fiabilité le profil pharmacocinétique de la substance étudiée chez un individu.

Le travail présenté dans ce manuscrit a pour objectif de déterminer un modèle de LSS précis et fiable capable de prédire l'exposition des nanocapsules lipidiques (NCL) de différents types dont la surface est soit nue, c'est-à-dire sans ajout de groupement chimiques (normales, Norm), soit pégylée (avec ajout de polyéthylène glycol, PEG), soit stearylée (avec ajout de stéarylamine, SA). Les NCL étudiées sont aussi de différentes tailles (85 nanomètres ou 50 nanomètres).

Un tel modèle doit permettre de prédire l'aire sous la courbe (AUC), c'est-à-dire le reflet de l'exposition avec un nombre de prélèvements inférieur ou égal à 3 dans le but de réduire considérablement le nombre de prélèvements à effectuer.

Ce travail a été réalisé en partie au sein du laboratoire MINT spécialisé dans les systèmes d'administration de médicaments à visée thérapeutique et dans l'optimisation des interactions de ces systèmes avec le vivant.

1. Nanomédicaments et nanocapsules

1.1. Définition²

Les nanomédicaments proviennent de l'application de la nanotechnologie aux sciences pharmaceutiques. Ce sont des structures de petites tailles de l'ordre du nanomètre capable de vectoriser le principe actif vers leurs sites d'action (Figure 1).

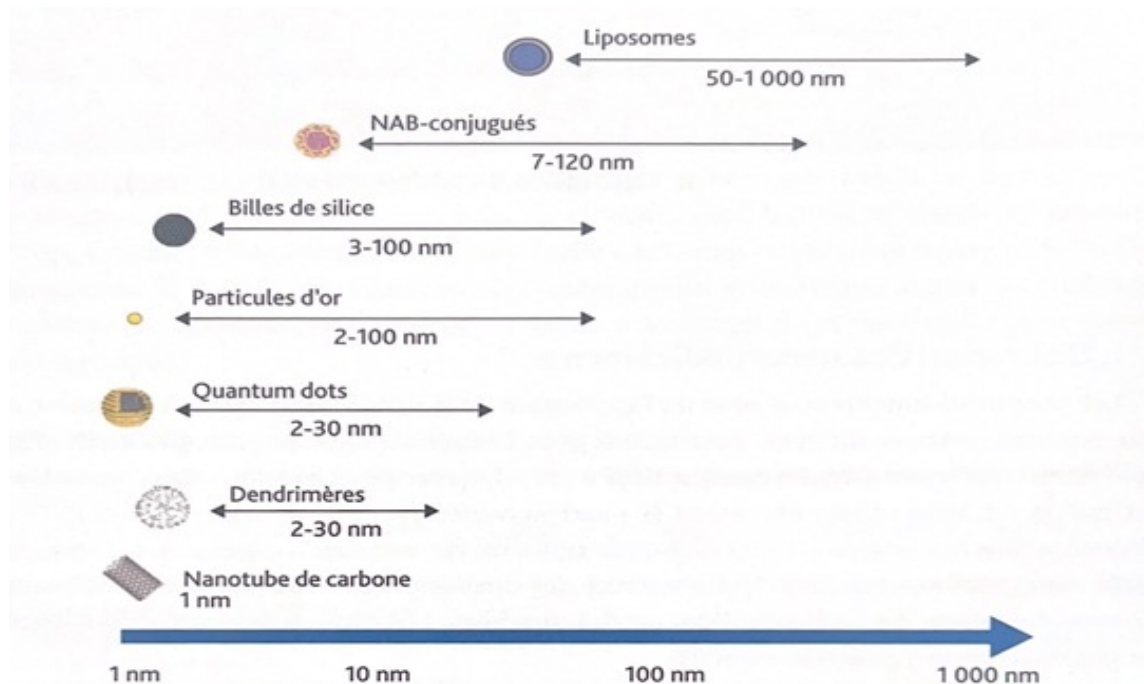


Figure 1 – Comparaison des tailles des nanomédicaments².

Ainsi, la pharmacocinétique du nanomédicament est principalement déterminée par la nanocapsule qui en est le vecteur plutôt que par le principe actif seul. De ce fait, les nanomédicaments ont plusieurs avantages et vont permettre la modification de la pharmacocinétique pour chacune des phases.

Premièrement, les nanomédicaments peuvent améliorer l'absorption des principes actifs. Ils favorisent également la solubilisation des principes actifs dans des milieux aqueux (milieux biologiques). De plus, ils peuvent permettre le franchissement de certaines barrières biologiques qui ne

seraient potentiellement pas accessibles directement aux principes actifs sans cette mise en forme galénique à l'échelle nanométrique.

Les nanomédicaments peuvent aussi être utilisés pour protéger le principe actif vis-à-vis d'une dégradation chimique ou enzymatique. En effet, certains principes actifs peuvent être très sensibles à l'environnement, comme le pH, la température ou encore à la présence de certaines enzymes par exemple. L'encapsulation peut donc présenter un avantage en protégeant le principe actif de son environnement.

La distribution du principe actif s'en retrouve également améliorée grâce à leur capacité de cibler des structures spécifiques. Elles permettent par ce biais de vectoriser le principe actif directement sur son site d'action limitant ainsi sa dispersion vers d'autres sites pour lesquels il aurait également une affinité. De ce fait, les nanomédicaments contribuent à réduire les effets indésirables des thérapies⁴.

Enfin, les nanomédicaments peuvent également être utilisés pour contrôler la libération des principes actifs. Cette libération est notamment contrôlée par les caractéristiques physico-chimiques de la nanoparticule permettant ainsi de moduler la pharmacocinétique des nanomédicaments⁵. Cet aspect sera détaillé plus loin.

Toutes ces propriétés montrent que les nanomédicaments sont des thérapies d'avenir et nécessitent des études approfondies afin de les décrire au mieux et de pouvoir approfondir leurs applications.

1.2. Classification des nanomédicaments

Il existe deux grandes classes de nanomédicaments : les formes encapsulées et les formes dites conjuguées.

Les formes encapsulées servent de réservoir pour le principe actif, elles jouent le rôle de vecteur pour l'acheminer vers sa cible. Parmi les principales formes de nanomédicaments encapsulées les dendrimères, les micelles, les liposomes ou encore les nanocapsules sont retrouvés. À l'heure actuelle, les formes encapsulées sont les plus nombreuses parmi les nanomédicaments.

Les formes conjuguées se définissent par l'association directe du principe actif à une macromolécule qui pourrait être un anticorps, un folate (comme la Vintafolide), une nanoalbumine (nanoparticule formée d'albumine) ou encore une chaîne pégylée comme le KADCYLA® (trastuzumab emtansine) par exemple. Cette liaison se fait *via* un *linker* qui sert de lien entre le principe actif et la macromolécule².

Les différentes catégories de nanomédicaments sont brièvement présentées dans la partie suivante.

1.2.1. Les formes encapsulées

1.2.1.1. Dendrimères

Les dendrimères sont des polymères sphériques en arbre composés de plusieurs couches autour d'un noyau central (Figure 2). Ce sont des structures amphiphiles c'est-à-dire qui comporte une partie hydrophobe et une partie hydrophile. Cette propriété facilite la solubilisation des médicaments hydrophobes en exploitant l'affinité de leur partie hydrophile pour le milieu aqueux^{6,7}.

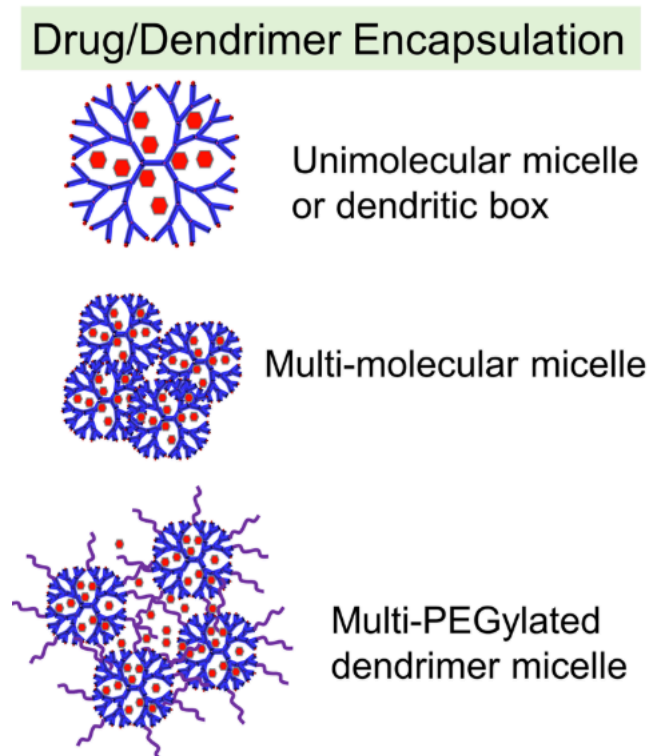


Figure 2 – Représentation de dendrimères encapsulant un principe actif⁷.

1.2.1.2. Micelles

Les micelles sont des structures sphériques qui se forment par l'auto-assemblage de polymères de nature amphiphile (Figure 3). Ce sont des structures ayant un cœur hydrophobe et une couronne hydrophile. Cependant, il existe aussi des micelles inverses ayant un cœur hydrophile et une couronne hydrophobe.

Les micelles peuvent également solubiliser des médicaments hydrophobes tout comme les dendrimères^{8,9}.

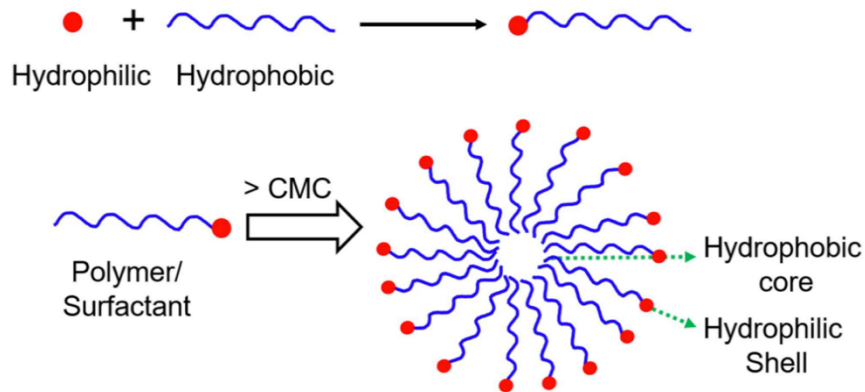


Figure 3 : Illustration de la formation d'une micelle⁸.

1.2.1.3. Liposomes

Les liposomes sont des vésicules sphériques composés de bicouches phospholipidiques. La particularité de ces particules est qu'elles présentent un noyau aqueux capable d'encapsuler des substances hydrophiles et une bicouche lipidique qui leur permet d'encapsuler des médicaments hydrophobes (Figure 4)^{10,11}.

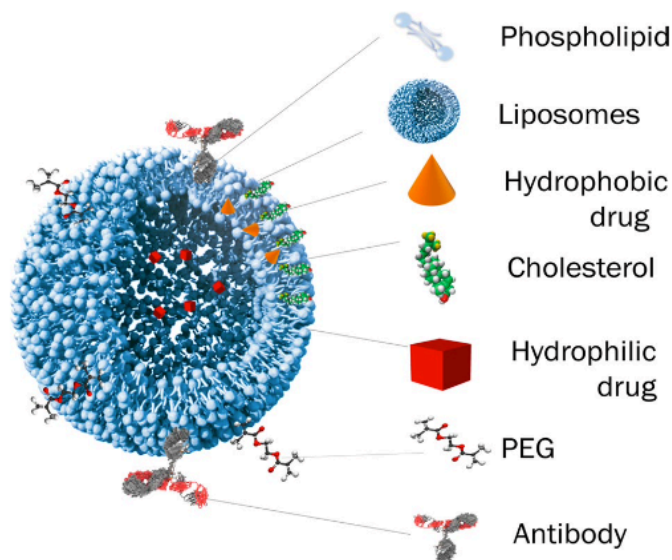


Figure 4 : Représentation d'un liposome¹¹.

1.2.1.4. Nanocapsules

Les nanocapsules sont des particules sphériques entourées d'une coque externe dont le noyau peut être solide ou liquide. Le noyau peut être constitué de lipides, de polymères ou d'huiles et il contiendra généralement le principe actif. La coque elle, peut être composée de polymères ou de lipides (Figure 5).

Les nanocapsules ont pour fonction principale de protéger le principe actif et de contrôler sa libération¹². Il existe un cas particulier de nanocapsules : les NCL. Ces dernières sont des nanocapsules lipidiques avec une couronne solide, composée de lécithines et de PEG et d'un noyau liquide formé de triglycérides¹³.

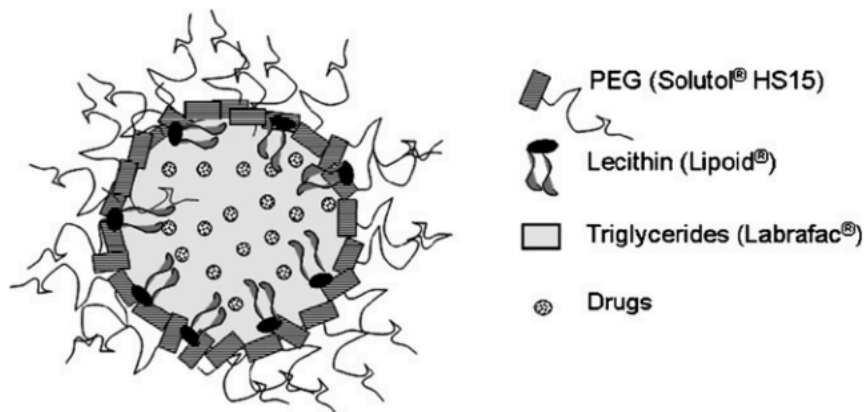


Figure 5 : Représentation schématique d'une nanocapsule lipidique (NCL)¹³.

1.2.2. Les formes conjuguées

1.2.2.1. Les anticorps conjugués aux principes actifs (*antibody drug conjugates*)

Les *antibody drug conjugates* (ADC) sont composés d'anticorps et d'un principe actif. Ces deux molécules sont reliées par un *linker* (lien) pour les solidariser afin que le déplacement de l'un entraîne le déplacement de l'autre (Figure 6). L'anticorps est spécifique d'un récepteur et permet d'acheminer le principe actif vers sa cible¹⁴.

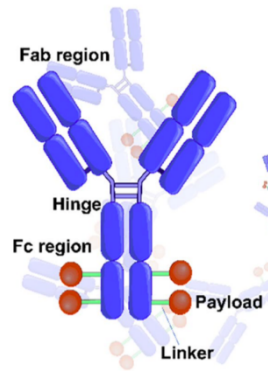


Figure 6 : Représentation d'un ADC¹⁴.

1.2.2.2. Médicaments liés à des nanocapsules d'albumine (nab-drugs)

Les *nab-drugs* exploitent l'albumine, une protéine endogène pour encapsuler des médicaments lipophiles via des interactions hydrophobes dans sa structure interne (Figure 7). En effet, l'utilisation de l'albumine est stratégique car elle permet d'améliorer la distribution de certains principes actifs¹⁵.

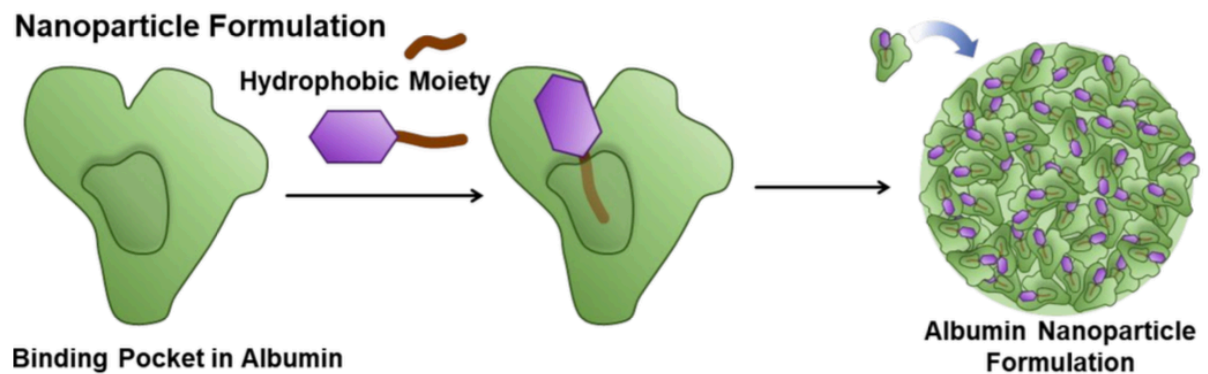


Figure 7 : Schéma illustrant la formation des nab-drugs¹⁵.

1.3. Applications médicales

Les nanomédicaments trouvent leur application dans de nombreux domaines notamment ceux qui nécessitent un ciblage spécifique. En effet, en dirigeant les principes actifs vers leur site d'action, la vectorisation contribue à limiter leurs effets indésirables.

Voici quelques applications médicales en cancérologie, cardiologie, infectiologie, neurologie ainsi que dans les maladies auto-immunes.

1.3.1. Cancérologie^{16,3}

L'utilisation des nanomédicaments en cancérologie permet d'améliorer la sélectivité, l'efficacité et la tolérance des traitements anticancéreux tout en diminuant les effets indésirables de ces thérapies souvent très toxiques. Ces propriétés de ciblage sont aussi bien utilisées en thérapeutique (chimiothérapie, radiothérapie, immunothérapie) qu'en diagnostic

(imagerie médicale). À titre d'exemple, l'Abraxane® (paclitaxel lié à l'albumine) est utilisé dans le traitement des cancers.

1.3.2. Cardiologie¹⁷

Les nanomédicaments sont utilisés dans le traitement des maladies cardiovasculaires. Ils peuvent par exemple être employés dans le traitement de l'athérosclérose ou encore dans le traitement de l'infarctus du myocarde. En effet, les nanomédicaments permettent de cibler spécifiquement les lésions athéromateuses, l'endothélium inflammé ainsi que les fibroblastes cardiaques activés ce qui améliore l'efficacité du traitement. Ils peuvent aussi être utilisés comme agent d'imagerie moléculaire afin de diagnostiquer et d'identifier les plaques athéromateuses. Enfin, ils ont aussi la capacité d'améliorer la régénération des tissus et la récupération fonctionnelle dans les troubles cardiovasculaires grâce à l'utilisation de nanoliposomes.

1.3.3. Infectiologie¹⁸

Ils ont également trouvé une application en infectiologie en améliorant le profil pharmacocinétique de certains médicaments antiviraux tout en diminuant la toxicité de ces derniers. De plus, certains nanomédicaments peuvent être utilisés comme agents virucides ou neutralisants augmentant ainsi l'efficacité des traitements. Il est également possible de créer des nano-vaccins.

1.3.4. Neurologie¹⁹

Dans le domaine de la neurologie, plus particulièrement dans les maladies neurodégénératives, les nanomédicaments trouvent des applications dans le traitement de la maladie d'Alzheimer et de Parkinson. Par exemple, les nanoparticules d'or ont pour effet d'inhiber l'agrégation des protéines A β 42 dans la maladie d'Alzheimer. Pour la maladie de Parkinson, les nanoparticules d'argent réduisent la neurotoxicité liée à l'inflammation cérébrale. Elles offrent également une meilleure biodisponibilité, une meilleure biocompatibilité et une cytotoxicité *in-vivo* réduite par rapport aux thérapies actuelles.

1.3.5. Maladies auto-immunes²⁰

Les nanomédicaments sont aussi utilisés dans le traitement des maladies auto-immunes. En effet, ils sont capables de garantir un ciblage spécifique des immunosuppresseurs vers leur site d'action. C'est le cas par exemple des nanoparticules en poly(lactic-co-glycolic acid) encapsulant du phosphate sodique de bétaméthasone. Ces nanoparticules augmentent l'efficacité de la substance active dans la polyarthrite rhumatoïde induite par un anticorps/adjuvant. Une injection de ces nanomédicaments contenant un tiers de la dose requise de bétaméthasone permettent d'obtenir des effets thérapeutiques similaires à ceux d'une injection de la dose entière requise de bétaméthasone.

2. La pharmacocinétique des nanomédicaments

2.1. Propriétés pharmacocinétiques des nanomédicaments

Les nanomédicaments vont suivre les différentes phases d'absorption, distribution, métabolisme et élimination comme des principes actifs non encapsulés (Figure 8). Cependant, la particularité de ces phases est qu'elles dépendent des propriétés du nanomédicament lui-même à travers ses caractéristiques physico-chimiques telles que la taille, la forme, la charge, la composition de la couronne, etc. Ces phases dépendent également de la pharmacocinétique du principe actif sous forme libre mais aussi des paramètres individuels du patient. Tous ces facteurs contribuent fortement à la complexité de la description de la pharmacocinétique des nanomédicaments².

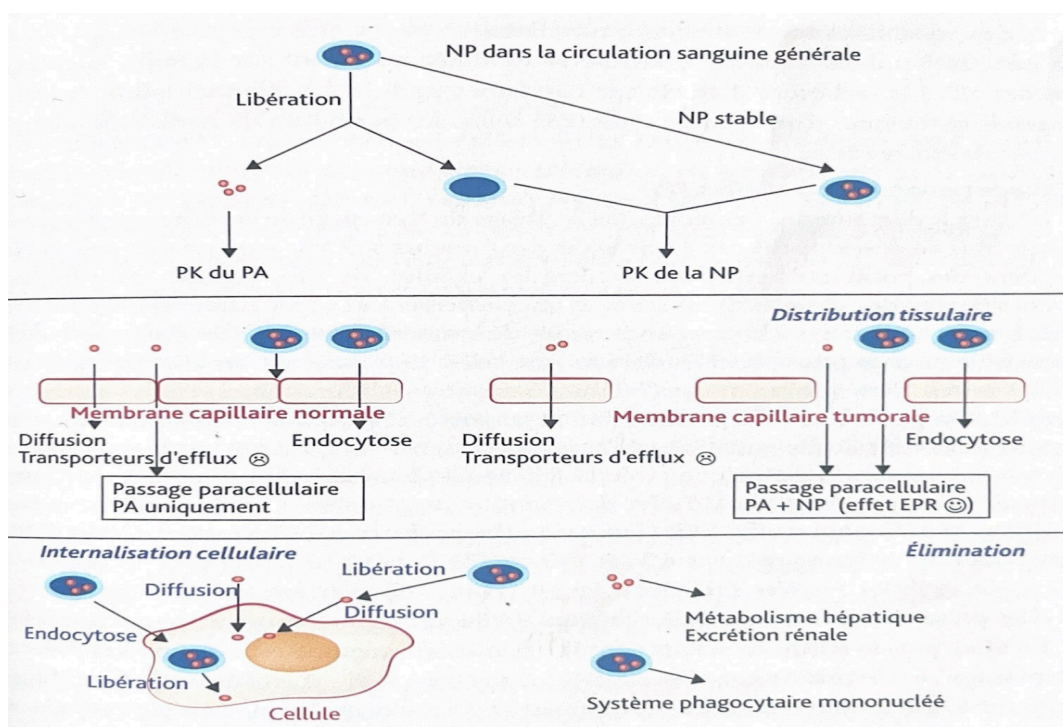


Figure 8 : Schéma illustrant la pharmacocinétique des nanomédicaments².

2.1.1. Absorption

Les nanomédicaments sont majoritairement administrés par voie intraveineuse. Cette voie d'administration permet de s'affranchir de la phase d'absorption car ils se retrouvent directement dans la circulation sanguine.

Dans le cas de la voie orale, les nanomédicaments peuvent être absorbés dans le tractus gastro-intestinal par différents mécanismes. Les nanomédicaments hydrophiles sont absorbés par endocytose (clathrines, cavéoles) ou par micropinocytose (absorption par de petites vésicules), deux processus connus comme étant des processus saturables et lents. Tandis que les nanomédicaments lipophiles sont principalement absorbés par transport transcellulaire. Lorsque l'un de ces processus d'absorption est

saturé ou bloqué, un relais existe entre ces derniers afin de maintenir l'absorption des nanomédicaments¹.

À noter que de multiples facteurs peuvent impacter l'absorption tel que les fluides gastriques, les fluides intestinaux qui au vu de leur pH, de la présence d'enzymes (lipases, peptidases) peuvent contribuer à la dégradation des nanomédicaments. D'autres facteurs modulent également l'absorption des nanomédicaments comme le mucus qui constitue une barrière physique à l'absorption, mais aussi chimiques à cause des interactions hydrophobes, électrostatiques, ou encore hydrogènes. Celles-ci représentent de réelles limites à la diffusion des nanomédicaments¹.

2.1.2. Distribution

Cette phase pharmacocinétique intervient après l'absorption et elle décrit la répartition du médicament dans les tissus de l'organisme.

Les nanomédicaments sont spécialement conçus pour moduler la distribution du principe actif dans les tissus. Ils ont été formulés spécifiquement pour optimiser cette phase afin qu'ils puissent atteindre plus spécifiquement les tissus cibles².

Après l'absorption, les nanomédicaments se retrouvent dans la circulation sanguine. Il a été démontré que très rapidement des protéines plasmatiques s'y fixent pour aboutir à la création d'une couronne modifiant ainsi la surface du nanomédicament (Figure 9). Par exemple, l'albumine présente en forte quantité dans le sang, a plutôt tendance à se lier à des nanomédicaments fonctionnalisés possédant des ligands neutres ou acides⁵.

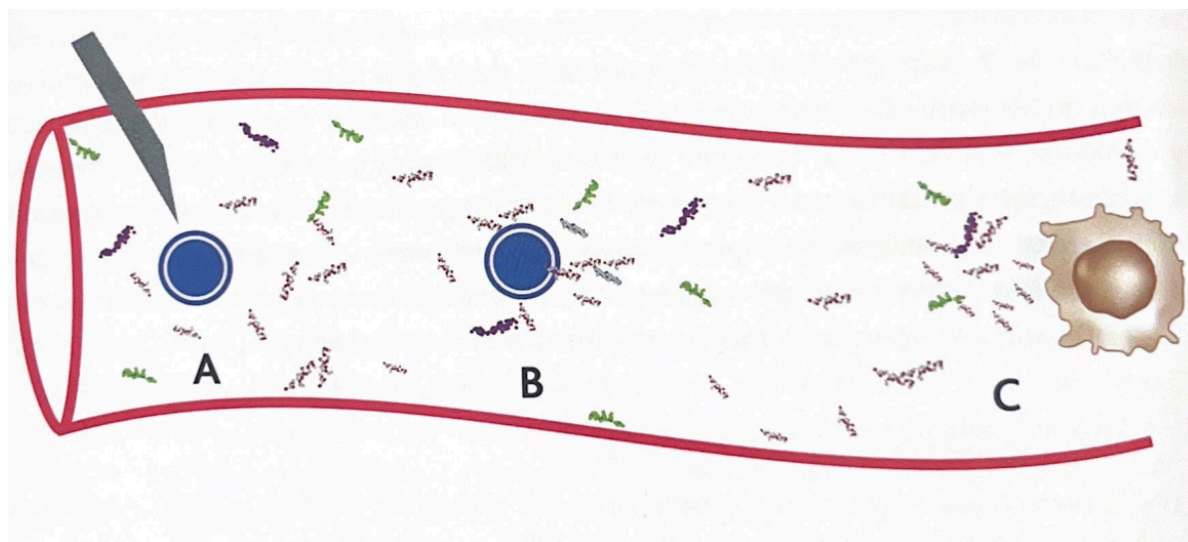


Figure 9 : Représentation de l'impact du phénomène d'opsonisation des nanomédicaments².

Sur la Figure 9, en A est représenté le phénomène d'opsonisation, en B le processus d'adsorption des protéines plasmatiques, en C la formation de la

couronne protéique et le ciblage des nanomédicaments par les macrophages².

Dans l'organisme, les nanomédicaments peuvent se distribuer à l'aide plusieurs mécanismes² tels que :

- L'effet *Enhanced Permeability and Retention* (EPR)
- Le ciblage actif

2.1.2.1. L'effet EPR

Le ciblage tumoral passif est une hypothèse qui a révolutionné le transport des molécules actives dans le domaine de l'oncologie. Ce ciblage s'effectue *via* l'effet EPR qui permet d'augmenter le temps de rétention et la quantité de principes actifs au sein de la tumeur. En effet, les tumeurs solides ont une angiogenèse souvent rapide et anarchique qui aboutit à la formation de vaisseaux endothéliaux présentant de nombreux pores et fenestrations. Ces fenestrations sont plus grandes que les fenestrations et pores des cellules saines. Cette différence de taille permet aux tumeurs solides d'acquérir plus de nutriments nécessaires pour que la tumeur puisse se développer. Cette propriété est exploitée pour faire passer des nanomédicaments qui vont cibler spécifiquement la tumeur. Ce phénomène se nomme l'effet de perméabilité augmentée ou EPR (Figure 10).

Cet effet est observé principalement pour les nanomédicaments qui ont une taille comprise entre 50 et 200 nm. Cependant pour pouvoir utiliser l'effet EPR, les nanomédicaments doivent être suffisamment stables, non agrégées et furtifs².

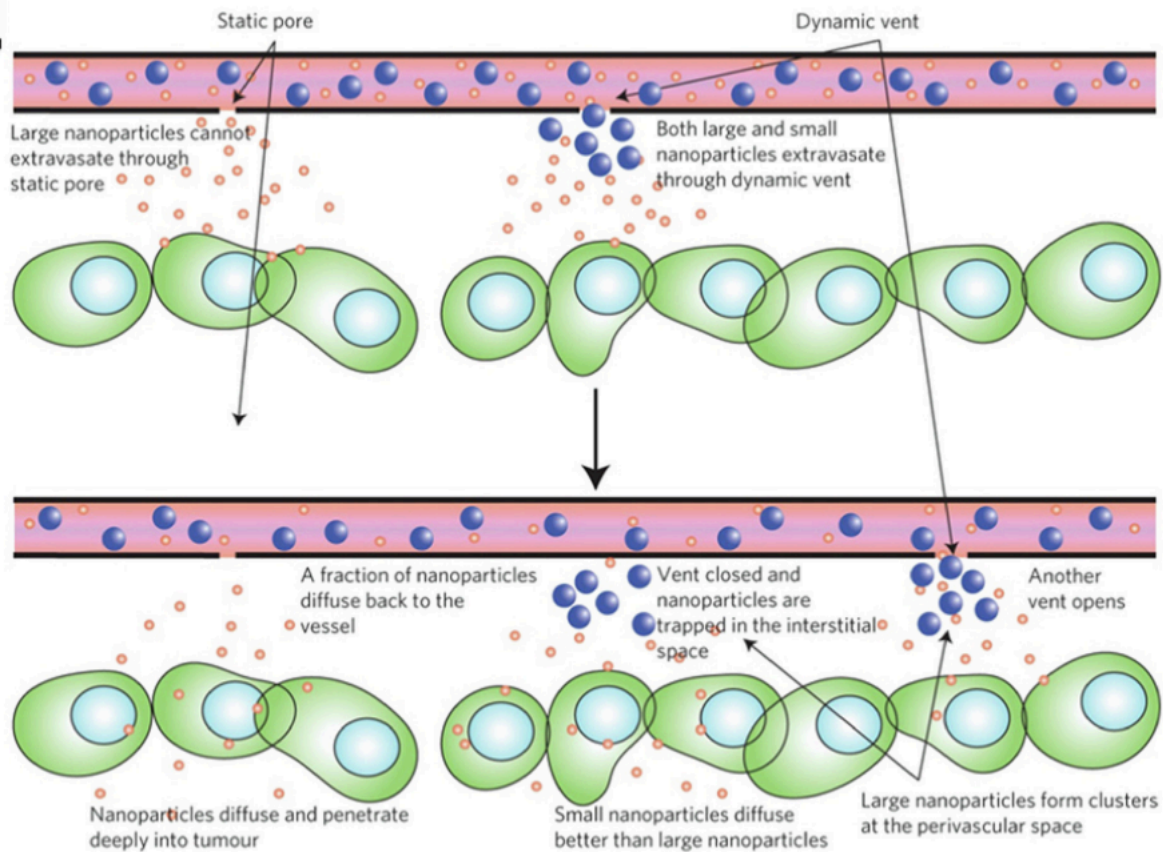
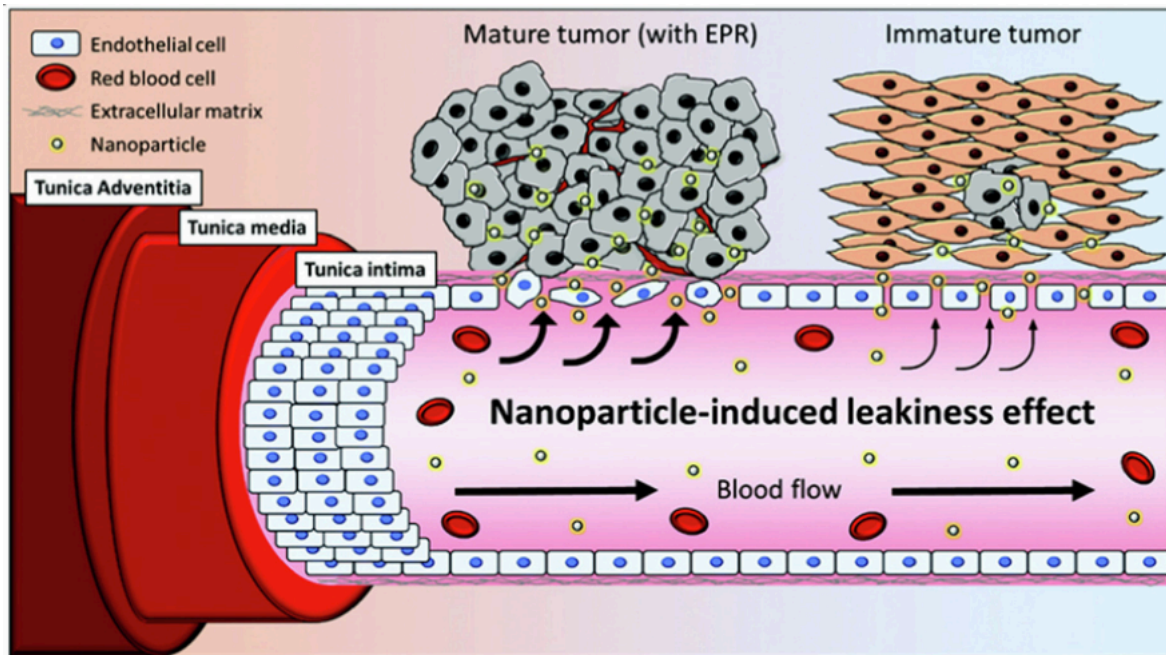


Figure 10 : Influence de l'effet EPR sur la distribution des nanomédicaments²².

2.1.2.2. Ciblage actif

Le ciblage actif utilise la fixation d'une molécule spécifique (ligand) et de forte affinité pour le tissu cible sur le nanomédicament. Comme l'effet EPR, le ciblage actif est principalement utilisé en cancérologie. Ce ligand va guider le ciblage du nanomédicament vers l'antigène tumoral pour lequel il est spécifique. Les anticorps sont les ligands les plus utilisés mais il en existe d'autres types comme les folates ou les nano-albumines. Une fois que le ligand interagit avec la protéine cible, le nanomédicament va pouvoir s'accumuler au sein de la tumeur. Cette interaction favorise la libération du principe actif à l'intérieur des cellules cibles par des mécanismes d'internalisation cellulaires ou à la suite de mécanismes de fusion de membrane².

2.1.2.3. Barrière à la distribution des nanomédicaments

Lorsque les nanomédicaments entrent dans la circulation, la première étape est la formation de complexes protéiques (opsonines qui se lient aux nanomédicaments) qui augmentent artificiellement la taille des nanomédicaments. Ces complexes protéiques se forment très rapidement (quelques dizaines de secondes après mise en contact avec le système sanguin). Elle a aussi pour conséquence de modifier la charge du nanomédicament². Cette augmentation de taille rend la pénétration des nanomédicaments plus difficile dans les différents tissus qui est taille-dépendante⁵. Ce phénomène se nomme l'opsonisation et aboutit à l'élimination des nanomédicaments grâce à une reconnaissance facilitée de ces derniers par le système immunitaire⁵.

Les protéines les plus impliquées dans ce phénomène sont l'albumine, les protéines du complément C3, les immunoglobulines, le fibrinogène ainsi que les facteurs de coagulation. Par ailleurs, le nombre de protéines s'adsorbant à la surface des nanomédicaments varie en fonction de leur charge, leur forme ou encore leur composition².

Au total ce phénomène d'opsonisation va impacter significativement la distribution des nanomédicaments qui parfois entraîne le masquage des ligands à la surface du nanomédicament empêchant donc la vectorisation et augmentant l'élimination des nanomédicaments *via* le SPM⁵.

Il a été démontré que seulement 0,3 à 3% des nanomédicaments parviennent à atteindre leur cible ce qui démontre l'importance de prendre en considération ces différentes limitations¹.

2.1.3. Métabolisme

Le lieu principal du métabolisme des nanomédicaments est le foie bien que cette phase puisse également avoir lieu dans la rate. Ce métabolisme s'effectue en deux phases. Premièrement, la phase de fonctionnalisation du nanomédicament qui permet l'ajout de groupements fonctionnels tout en augmentant sa polarité^{5,24}. Divers cytochromes peuvent être impliqués dans cette phase première de fonctionnalisation, le cytochrome P450 joue un rôle important dans cette phase en catalysant de nombreuses réactions

d'oxydation, de réduction ainsi que des réactions d'hydrolyse permettant de préparer les nanomédicaments à la phase de conjugaison¹.

Ensuite intervient la phase de conjugaison^{5,25}. Cette phase de conjugaison permet d'augmenter l'hydrophilie du nanomédicament et de ses métabolites pour favoriser son élimination²⁴.

2.1.4. Élimination

Les nanomédicaments ont pour objectif de limiter l'élimination des principes actifs encapsulés dans ces derniers². Néanmoins, les nanomédicaments passent tout de même par une phase d'élimination plus ou moins importante. Divers mécanismes sont mis en causes dans la phase d'élimination (Figure 12) :

- L'instabilité dans les fluides biologiques
- L'élimination hépatobiliaire
- L'élimination rénale
- L'élimination par le système lymphatique

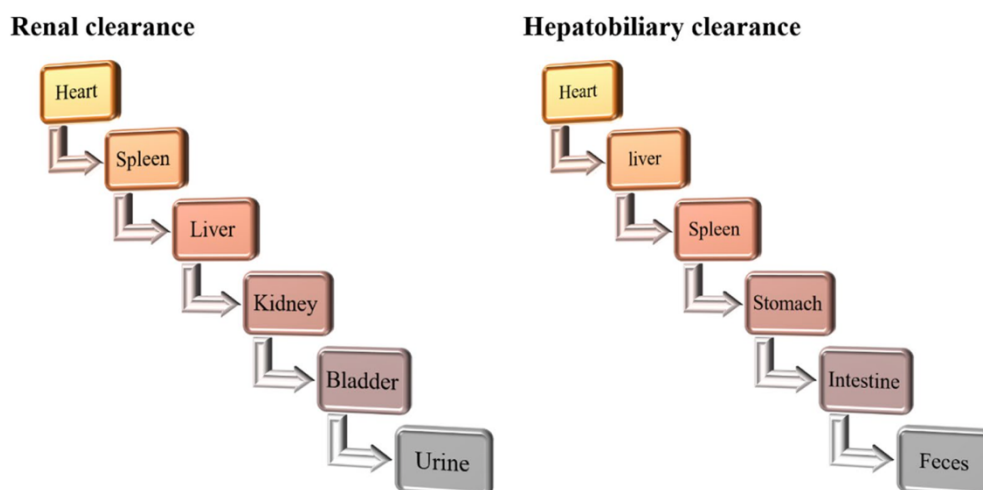


Figure 12 : Les différentes voies d'élimination des nanomédicaments⁵.

2.1.4.1. L'instabilité dans les fluides biologiques

Les nanomédicaments présentent un rapport surface-volume élevé ce qui en fait des substances très réactives. En effet, plus les nanomédicaments sont petits, plus la tension de surface est grande ce qui favorise la coalescence et l'agglomération des nanomédicaments. Ces deux phénomènes entraînent alors une instabilité de ces derniers. De plus, la circulation sanguine est composée de plusieurs protéines susceptibles d'interagir avec les nanomédicaments. Ces interactions peuvent avoir pour effet de déstabiliser les nanomédicaments et donc de libérer le principe actif avant qu'il n'arrive sur son site d'action².

2.1.4.2. L'élimination hépatobiliaire

Les hépatocytes et les cellules de Kupffer (macrophages hépatiques) sont impliqués dans l'élimination hépatobiliaire mais de façon différente (Figure 13). Les hépatocytes agissent plutôt dans l'excrétion par la bile tandis que les cellules de Kupffer jouent un rôle majeur dans la dégradation intracellulaire pour éliminer les nanomédicaments *via* le SPM²¹.

Les nanomédicaments opsonisés et phagocytés par les macrophages vont être dégradés majoritairement dans le foie et la rate. Les produits de la dégradation sont ensuite absorbés par les macrophages ou rejetés dans le secteur extracellulaire. Les nanomédicaments non dégradés restent ainsi piégés jusqu'à la mort du macrophage dans les phagosomes². Aussi, les nanomédicaments peuvent être éliminés par la bile après avoir été dégradés et transportés par les hépatocytes, cette élimination s'opère ensuite *via* les fèces⁵.

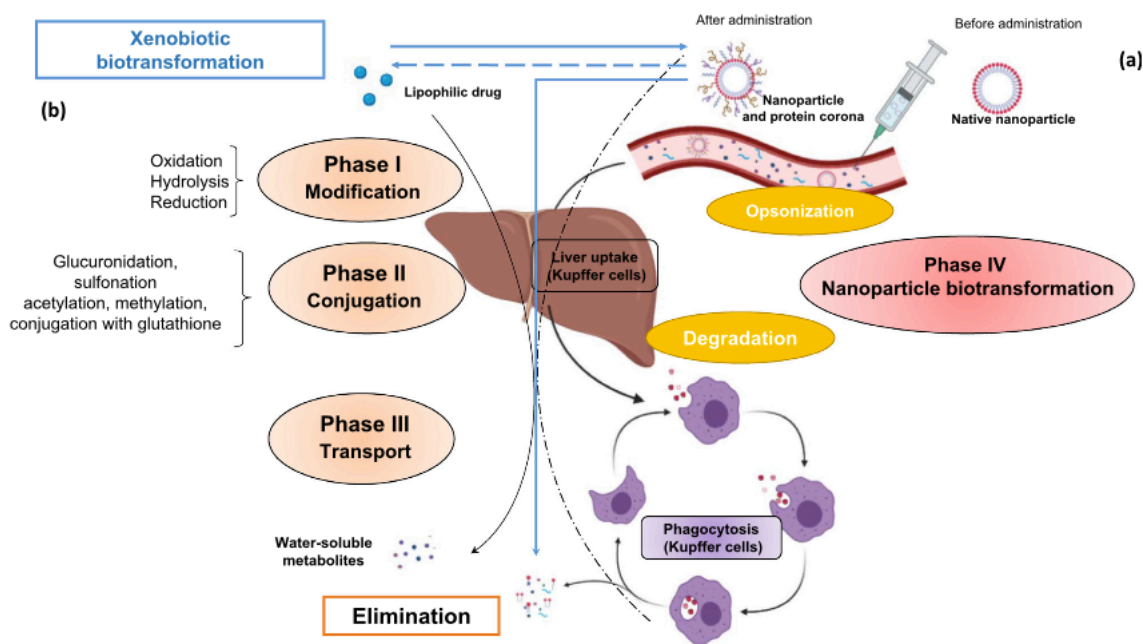


Figure 13 : Schéma illustrant la dégradation et l'élimination des nanomédicaments par le SPM¹.

2.1.4.3. Élimination rénale

L'élimination rénale des nanomédicaments peut se faire par filtration glomérulaire *via* les fenestrations de l'endothélium du glomérule². Les nanomédicaments interagissent alors avec les capillaires au niveau du glomérule en les traversant ce qui leur permet d'entrer dans la capsule de Bowman avant d'être réabsorbés par le tubule rénal. Certains nanomédicaments sont renvoyés dans la circulation sanguine tandis qu'une autre partie est excrétée dans les urines (Figure 14)⁵.

La limite de filtration glomérulaire est estimée à environ 65 kilodaltons (kDa) soit le poids moléculaire de l'albumine ce qui correspond à une taille comprise entre 6 et 8 nm. Ainsi, seuls les nanomédicaments ayant une taille inférieure à 6 nm sont filtrés librement par le glomérule¹. Les nanomédicaments ayant une taille de l'ordre de 4-5 nm sont eux filtrés beaucoup plus rapidement car cette taille correspond à la taille des fenestrations endothéliales, de la membrane basale du glomérule ainsi qu'à celle des podocytes.

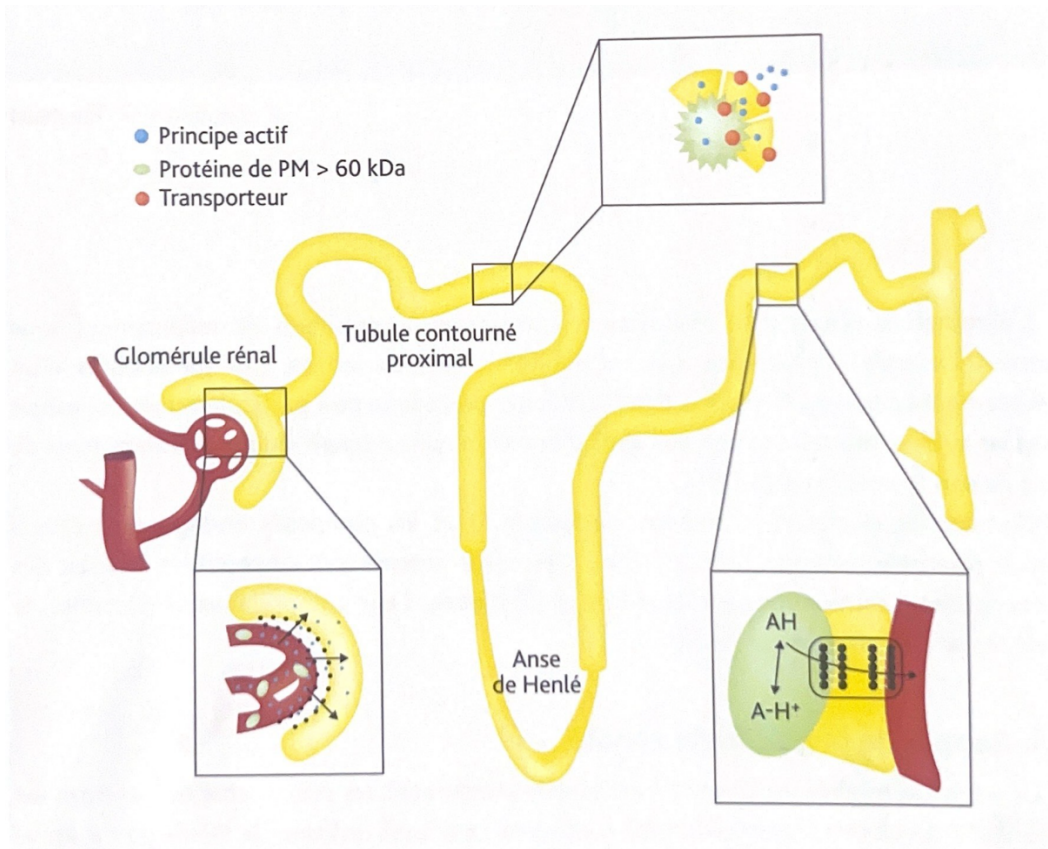


Figure 14 : Représentation de la filtration glomérulaire, de la sécrétion tubulaire et de la réabsorption tubulaire des médicaments²⁶.

2.1.4.4. L'élimination par le système lymphatique

Le système lymphatique (Figure 11) joue un rôle important dans l'élimination des nanomédicaments. En effet, les capillaires et le flux lymphatique ont leur importance car ce système permet de véhiculer les nanomédicaments de l'interstitium vers le système lymphatique et les ganglions lymphatiques afin de faciliter leur élimination.

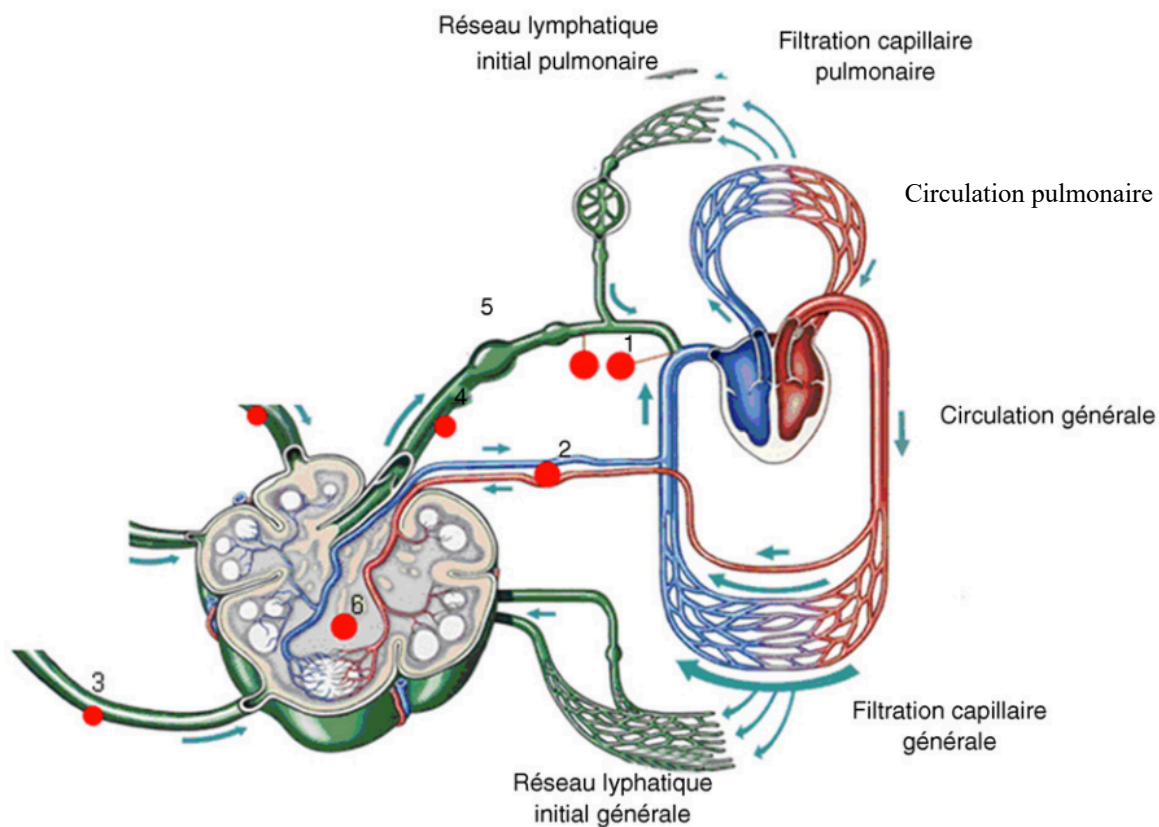


Figure 11 : Schéma illustrant le système cardiovasculaire et le système lymphatique²³.

2.1.5. Influence des propriétés physico-chimiques sur leur pharmacocinétique des nanomédicaments

2.1.5.1. Taille

La taille des nanomédicaments est un paramètre important car elle impacte plusieurs phases pharmacocinétiques (Figure 16).

Premièrement, la taille a un impact sur l'adsorption des protéines à la surface des nanomédicaments. Une étude a montré qu'après injection de nanomédicaments constitués de polyhexadéylcyanoacrylate (PHDCA) et pégylés de taille différente, les nanomédicaments de petites tailles (de l'ordre de 80 nm) restent deux fois plus longtemps dans la circulation sanguine que les gros nanomédicaments (de plus de 200 nanomètres). Ce qui signifie que les nanomédicaments de plus de 200 nm ont tendance à être éliminés plus rapidement que les nanomédicaments qui ont une taille

de l'ordre de 80 nm. Ce constat pourrait s'expliquer par le fait que les nanoparticules de petites taille pourraient potentiellement moins adsorber d'opsonines sur leur surface à cause de leur faible taille leur permettant potentiellement d'échapper à l'élimination par le SPM⁵.

Elle conditionne également sa capacité à pouvoir bénéficier de l'effet EPR, D'après la littérature, la taille optimale est comprise entre 50 et 100 nm voire jusqu'à 300 nm^{2,5}.

La taille des nanomédicaments conditionne également le mécanisme d'internalisation de ces derniers qui seront mis en jeu (Figure 15). En effet, les nanomédicaments de taille importante (un micromètre) subiront la phagocytose. Les nanomédicaments ayant une taille comprise entre quelques centaines de nanomètres et quelques micromètres subiront la pinocytose et la macropinocytose. Les nanomédicaments ayant une taille inférieure à 200 nm seront internalisés par endocytose impliquant les clathrines ou les cavéolines².

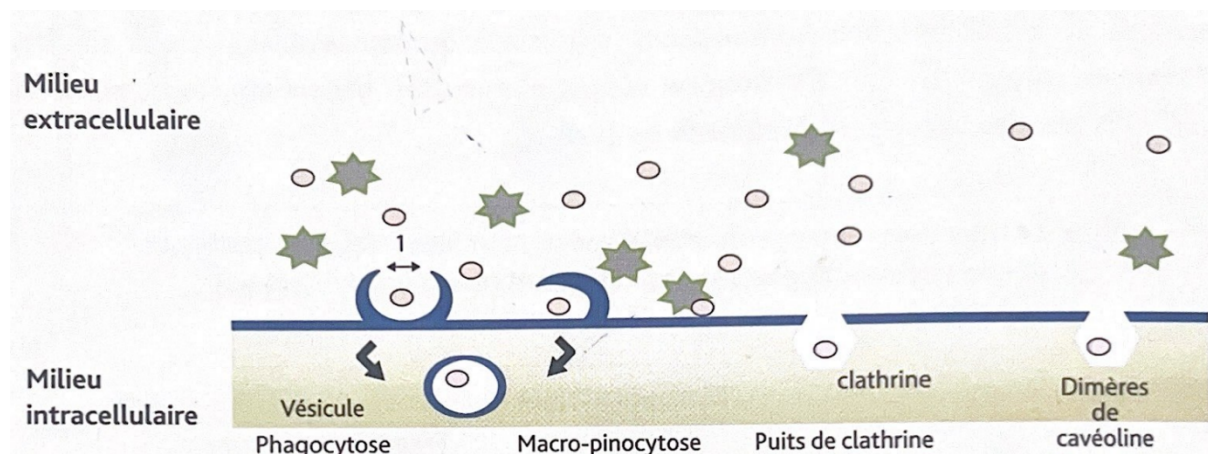


Figure 15 : Les différents modes d'internalisation des nanomédicaments².

La taille des nanomédicaments conditionne aussi l'élimination rénale de ces derniers. Par ailleurs, les nanomédicaments ayant une taille supérieure ou égale à 200 nm ont plutôt tendance à activer le système lymphatique et sont donc éliminés plus rapidement.

La taille a également un impact sur le temps de demi-vie. En effet, il a été constaté que le temps de demi-vie est plus long pour les nanomédicaments de 100 nm que pour les nanomédicaments de 10 nm car les nanomédicaments de 100 nm subissent moins l'élimination par le rein⁵.

À la suite de l'analyse des données issues de la littérature, il semblerait que la taille idéale des nanomédicaments soit de 100 nm afin d'avoir une bonne distribution de ces derniers et une clairance d'élimination totale maîtrisée⁵.

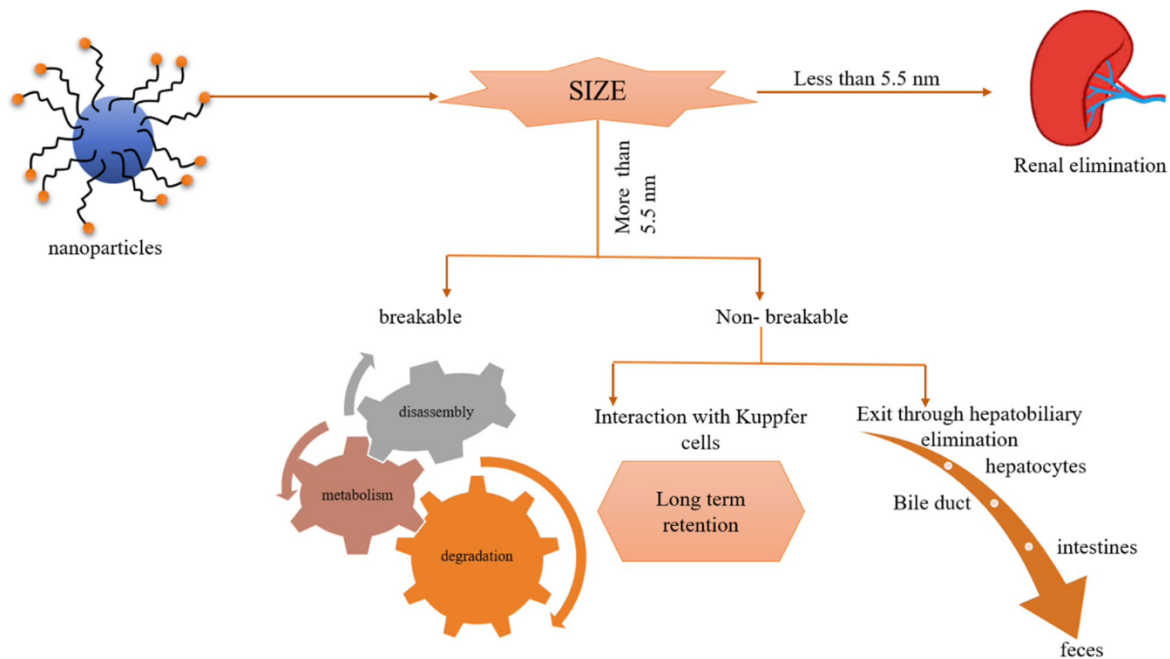


Figure 16 : Impact de la taille sur la pharmacocinétique des nanomédicaments⁵.

2.1.5.2. Composition

La pharmacocinétique des nanomédicaments est fortement influencée par leur composition. En effet, la composition influence la stabilité du nanomédicament ainsi que sa capacité à libérer le principe actif. Le ratio de cholestérol et l'organisation des lipides sont notamment des paramètres déterminants pour la stabilité et la libération du principe actif².

De plus, la surface du nanomédicament est aussi un paramètre qui impacte la pharmacocinétique des nanomédicaments. En effet, elle est le reflet de la solubilité du nanomédicament, de son potentiel d'agrégation, de sa capacité à traverser les barrières biologiques, de sa bio-compatibilité mais aussi de ses propriétés de ciblage. Il a été démontré que l'ajout de molécule de polyéthylène glycol à la surface (pégylation), permet de réduire la captation du nanomédicament par certains organes permettant ainsi aux nanomédicaments pégylés d'échapper au système immunitaire². La plupart du temps les nanomédicaments ont une surface hydrophile pour interagir avec les systèmes biologiques⁵.

Enfin, un autre point qui mérite de l'attention et qui peut impacter la pharmacocinétique des nanomédicaments est l'homogénéité de la formulation. En effet, il est possible de retrouver une population de nanomédicaments différente au sein d'une même formulation par exemple des nanomédicaments sous forme de nanocapsules coexistant avec des nanomédicaments sous forme de micelles. La pharmacocinétique de ce mélange doit prendre en compte la pharmacocinétique de chaque élément constitutif de cette population¹.

2.1.5.3. Charge électrique

La charge électrique apparente de surface également appelée potentiel zêta est un élément important à prendre en compte dans l'étude de la pharmacocinétique des nanomédicaments.

En effet, les nanomédicaments chargés positivement sont davantage absorbés par endocytose tandis que les nanomédicaments chargés négativement le sont moins que les nanomédicaments neutres. De plus, les nanomédicaments chargés négativement ont tendance à utiliser des voies d'internalisation non spécifiques⁵.

Cette charge influence la stabilité et le potentiel d'agrégation des nanomédicaments avec les protéines dans la circulation sanguine, ce qui peut impacter la demi-vie. Les nanocapsules cationiques ont tendance à s'agréger avec les protéines du plasma qui ont un point isoélectrique inférieur à 5 tandis que les nanocapsules anioniques ont tendance à s'agréger avec les protéines du plasma qui ont un point isoélectrique supérieur à 5⁵.

Il a été démontré que cette charge impacte la clairance des nanomédicaments et plus particulièrement les nanomédicaments dont la nanocapsule est chargée positivement qui ont une clairance plus rapide ainsi qu'une accumulation plus importante au niveau de la tumeur que les nanomédicaments dont la nanocapsule est chargée négativement². Ce constat peut être expliqué par le fait que leur charge positive leur permettrait d'interagir avec les surfaces cellulaires chargées négativement⁵.

2.1.5.4. Forme

Les nanomédicaments peuvent avoir des formes diverses (Figure 17). En effet, ils peuvent être cubiques, sphériques ou encore cylindriques. Il a été démontré que les nanomédicaments sphériques pénètrent mieux les cellules mais plus difficilement les fenestrations intracellulaires². Aussi, Les nanomédicaments non sphériques circulent plus lentement, subissent moins la phagocytose par les macrophages et sont absorbés plus facilement par les cellules que les nanomédicaments sphériques. Une étude a également montré que les nanomédicaments discoïdaux sont ceux qui subissent le moins l'élimination hépatique. Une théorie pour expliquer ce phénomène est le fait qu'il y a une part importante de nanomédicaments discoïdaux qui se dirigent vers la paroi des vaisseaux ce qui facilite leur distribution dans les autres organes. Elle s'expliquerait aussi par le fait que les nanomédicaments discoïdaux échappent plus facilement à la phagocytose⁵.

De plus, plus la forme de la nanocapsule est cylindrique ou cubique plus l'élimination est ralentie donc plus le temps de demi-vie de la nanocapsule est augmenté². Enfin, certains organes ont des affinités plus prononcées pour certains types de nanocapsules. Par exemple, les nanocapsules avec une forme irrégulière ont plutôt tendance à s'accumuler dans la rate tandis

que les nanocapsules ayant une forme de tige s'accumule préférentiellement dans les poumons⁵.

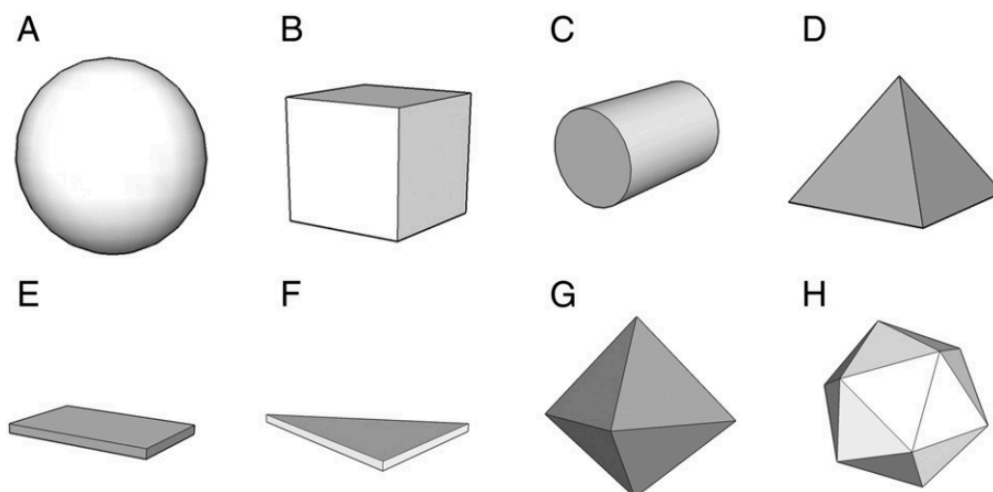


Figure 17 : Différentes formes de nanocapsules²⁷.

L'image A est une représentation schématique d'une nanocapsule sphérique, B une nanocapsule cubique, C une nanocapsule en bâtonnet, D une nanocapsule pyramidale, E une nanocapsule nanoplate, F une nanocapsule nanoplate triangulaire, G une nanocapsule octaédrique, H une nanocapsule icosaédrique.

2.1.6. Influence des variabilités interindividuelles sur la pharmacocinétique des nanomédicaments

Les propriétés pharmacocinétiques des nanomédicaments ne sont pas uniquement dépendantes de leurs caractéristiques physico-chimiques. Elles dépendent également des individus à qui ils sont administrés. Plusieurs paramètres sont à prendre en compte à savoir :

- L'âge
- Le poids
- La variabilité physiopathologique
- Le système immunitaire
- La dose
- La voie d'administration

2.1.6.1. L'âge

L'âge est une source de modification de la pharmacocinétique des nanomédicaments. En effet, les propriétés physiologiques d'un sujet jeune ne sont pas les mêmes que celles d'un sujet âgé. Il y a de grandes variations notamment en termes de maturité mais aussi en termes de croissance des systèmes impliqués dans les phases pharmacocinétiques. C'est notamment le cas pour les systèmes impliqués dans le métabolisme hépatique, dans l'élimination rénale, dans la résorption digestive mais aussi dans la répartition de la masse maigre et de la masse grasse. Ainsi, les capacités d'absorption de distribution, de métabolisme et d'élimination diffèrent en

fonction de l'âge du sujet. Ces différences sont à l'origine de modifications des profils pharmacocinétiques et pourraient être à l'origine d'une adaptation posologique en fonction de l'âge du sujet dans certains cas²⁸.

2.1.6.2. Le poids

Le poids est aussi un paramètre clé qui impacte les différentes phases pharmacocinétiques. C'est un paramètre qui permet de calculer la dose à administrer pour de nombreux médicaments. Cette dose a également un impact sur le profil pharmacocinétique.

Les principaux facteurs qui expliquent l'impact du poids en pharmacocinétique incluent les modifications de la répartition masse maigre/masse grasse, les modifications de la fonction cardiovasculaire, de la fonction rénale, de la fonction hépatique mais également les modifications de la quantité de protéines inflammatoires et d'adipokines²⁹.

2.1.6.3. Variabilité physio-pathologique

La pharmacocinétique des nanomédicaments est aussi impactée par l'évolution des pathologies. Dans le cas des tumeurs, l'effet EPR est différent en fonction du type tumoral mais aussi d'une tumeur à une autre. Cette différence impacte donc la pharmacocinétique des nanomédicaments car l'effet EPR est impliqué dans la phase de distribution intra-tumorale des nanomédicaments. De plus, les tumeurs sont des pathologies évolutives et nécessiteraient dans l'idéal une constante adaptation de la formulation des nanomédicaments².

2.1.6.4. Importance du système immunitaire²

Chaque patient a un système immunitaire qui lui est propre et qui peut varier en fonction de certaines pathologies. En effet, certaines pathologies comme les maladies inflammatoires ou les maladies auto-immunes altèrent le fonctionnement du système immunitaire. De plus, certains patients suivent des thérapies comme les traitements myélotoxiques qui affectent leur système immunitaire.

Les nanomédicaments administrés chez ces patients n'auront donc pas la même pharmacocinétique que chez des patients immuno-compétents. En effet, l'élimination des nanomédicaments se fait grâce au système SPM faisant intervenir le système immunitaire. Ceci implique qu'un système immunitaire déficient diminue la clairance des nanomédicaments et par conséquent pourrait entraîner potentiellement un surdosage et donc une toxicité du nanomédicament. À l'inverse, un système immunitaire suractivé augmente la clairance des nanomédicaments ce qui pourrait provoquer un sous-dosage et donc une absence ou une diminution de l'effet thérapeutique du nanomédicament².

2.2. Paramètres et modèles pharmacocinétiques

2.2.1. Paramètres pharmacocinétiques

Les paramètres pharmacocinétiques sont des paramètres qui permettent l'interprétation fine et précise du devenir d'une substance après administration. Il existe deux catégories de paramètres pharmacocinétiques³⁰ :

- Les paramètres graphiques : ils peuvent être représentés sur les courbes représentant l'évolution des concentrations au cours du temps. Les principaux paramètres graphiques sont le temps de demi-vie, le temps de résidence moyen ou encore l'aire sous la courbe en fonction du temps (AUC). Ce groupe de paramètres sont dits secondaires car ils sont utilisés pour la description de la pharmacocinétique.
- Les paramètres mécanistiques appelés aussi paramètres primaires ou fondamentaux sont des paramètres calculés à partir des paramètres secondaires qui définissent les propriétés pharmacocinétiques de la molécule d'intérêt. Ces paramètres sont composés essentiellement de la constante d'élimination (k_e), du volume de distribution (V_d) et de la clairance d'élimination totale (CL).

D'autres paramètres graphiques et mécanistiques existent³⁰ mais ne feront pas l'objet d'étude approfondie. En effet, les paramètres évoqués ci-dessus sont suffisants pour comprendre la suite de ce manuscrit.

2.2.1.1. Paramètres graphiques

2.2.1.1.1. Aire sous la courbe (AUC)³¹

L'AUC est un paramètre qui reflète l'exposition de l'organisme à la substance étudiée. Elle représente approximativement la concentration moyenne en médicament à un facteur près qui est la durée d'exposition. Mathématiquement, l'AUC se définit comme l'aire sous la courbe des concentrations en fonction du temps.

L'utilisation de l' AUC_{inf} qui calcule l'aire sous la courbe de 0 jusqu'à l'infini est préférable pour calculer l'AUC de manière précise. Voici le calcul de l'AUC, avec $C(t)$ la fonction qui décrit l'évolution des concentrations en fonction du temps et les constantes Z (ayant l'unité d'une concentration) et δ qui permettent de décrire cette évolution :

$$\begin{aligned}AUC_{inf} &= \int_0^{\infty} C(t).dt \\AUC_{inf} &= \int_0^{\infty} Ze^{-\delta.t} dt = \lim_{a \rightarrow \infty} \int_0^a Ze^{-\delta.t} dt \\AUC_{inf} &= \lim_{a \rightarrow \infty} \left[Z \times \frac{e^{-\delta.t}}{-\delta} \right]_0^a \\AUC_{inf} &= \lim_{a \rightarrow \infty} \left[-\frac{Z}{\delta} \times (e^{-\delta.a} - e^{-\delta.0}) \right] \\AUC_{inf} &= -\frac{Z}{\delta} \times (0 - 1) = \frac{Z}{\delta}\end{aligned}$$

Lorsque $C(t)$ est une somme de n exponentielle, l' AUC_{inf} devient :

$$AUC_{inf} = \sum_{i=1}^n \frac{Z_i}{\delta_i}$$

En réalité cette AUC ne représente pas la réalité car il est impossible d'obtenir les valeurs des concentrations jusqu'à un temps infini. Cette équation est extrapolée et peut se décomposer de la manière suivante :

$$AUC_{inf} = AUC_{last} + AUC_{extrapolée}$$

L' AUC_{last} correspond à la partie calculable à partir des observations réelles, c'est la partie non extrapolée. Tandis que l' $AUC_{extrapolée}$ est la partie de l'AUC calculée entre le dernier point (last) et l'infini, c'est la zone dont

l'information n'est pas connue. L' AUC_{last} se calcule à l'aide de la méthode des trapèzes qui repose sur l'idée que l'aire sous le curve peut être divisée en plusieurs trapèzes dont la somme des aires est égale à l' AUC_{last} .

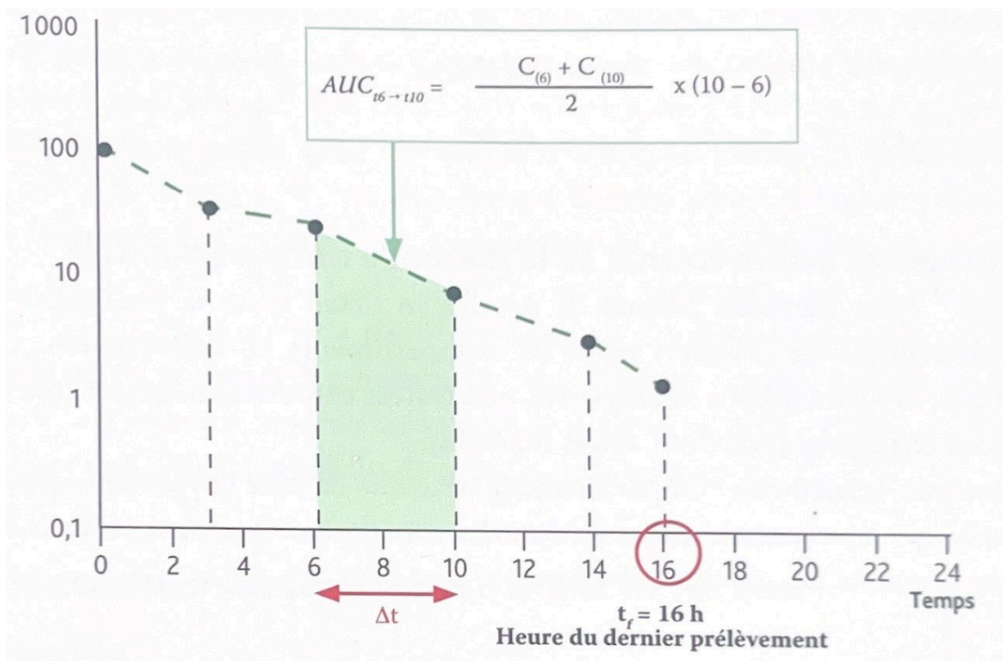


Figure 18 : Exemple de calcul de l'aire d'un trapèze pour calculer l'AUC entre deux points³¹.

La méthode des trapèzes (Figure 18) permet donc de calculer l' AUC_{last} à l'aide de cette formule :

$$AUC_{last} = \sum_{i=1}^{n-1} \frac{(C_i + C_{i+1})}{2} \times (t_{i+1} - t_i)$$

L' $AUC_{extrapolée}$ peut être calculée à l'aide d'une intégrale en posant $Z = C_f$ qui est la dernière concentration observée :

$$AUC_{extrapolée} = \int_{t_f}^{\infty} Z e^{-\delta \cdot t} dt = \frac{C_f}{\delta}$$

Ces AUCs peuvent être calculées à l'aide d'un logiciel spécifique comme PKanalix® (Lixoft, Paris France). Cela permet de calculer les différents paramètres pharmacocinétiques en utilisant des méthodes spécifiques dédiées à savoir la méthode *linear trapezoidal linear* (calcul linéaire de l'AUC et interpolation linéaire), la méthode *linear-log trapezoidal* (calcul linéaire de l'AUC et interpolation linéaire avant le C_{max} , calcul logarithmique de l'AUC et interpolation logarithmique après le C_{max}), la méthode *linear up log down* (calcul linéaire de l'AUC et interpolation linéaire si la concentration augmente et un calcul logarithmique et interpolation logarithmique si elle diminue) ainsi que la méthode *linear trapezoidal linear-log* (calcul linéaire

de l'AUC, interpolation linéaire avant le C_{\max} et calcul logarithmique après le C_{\max}). Ces méthodes ont pour avantage de calculer précisément l'AUC en fonction de la forme de la courbe des concentrations.

2.2.1.1.2. Le temps de demi-vie d'élimination³¹

Le temps de demi-vie d'élimination ($t_{1/2}$) est le temps au bout duquel la concentration plasmatique en médicament a diminué de moitié. L'équation qui décrit l'évolution des concentrations en fonction du temps est la suivante :

$$C(t) = C_0 \cdot e^{-k_e \cdot t}$$

À partir de la définition précédente, il est aisément possible de calculer le $t_{1/2}$:

$$\frac{C_A}{2} = \frac{C_0 \cdot e^{-k_e \cdot t_A}}{2} = C_0 \cdot e^{-k_e \cdot \left(\frac{t_1+t_A}{2}\right)}$$

Après simplification :

$$\frac{e^{-k_e \cdot t_A}}{2} = e^{-k_e \cdot \left(\frac{t_1+t_A}{2}\right)}$$

$$\frac{1}{2} = e^{-k_e \cdot \frac{t_1-t_A}{2}}$$

$$-\ln(2) = -k_e \cdot \frac{t_1}{2}$$

$$t_{1/2} = \frac{\ln(2)}{k_e}$$

2.2.1.1.3. Le temps de résidence moyen (MRT)³¹

Le temps de résidence moyen est un paramètre qui représente la durée de séjour moyenne d'une substance dans l'organisme. C'est la durée moyenne de l'ensemble des phases pharmacocinétiques (phases ADME). Le MRT se calcule de la façon suivante :

$$MRT = \frac{\int_0^{\infty} C(t) \cdot t \cdot dt}{\int_0^{\infty} C(t) \cdot dt} = \frac{AUMC_{inf}}{AUC_{inf}}$$

Dans cette équation un autre paramètre graphique est pris en compte, c'est l'AUMC. C'est un paramètre dont l'unité est une concentration multipliée par une unité de temps au carré. L'AUMC représente l'exposition globale du médicament dans l'organisme en fonction du temps.

2.2.1.2. Paramètres mécanistiques

2.2.1.2.1. La clairance d'élimination totale³²

La clairance d'élimination totale ou clairance d'élimination plasmatique est le facteur de proportionnalité entre la vitesse d'élimination du médicament et la concentration plasmatique du médicament. Cette clairance est un débit et peut s'exprimer par exemple en L/h. Son interprétation pharmacocinétique représente les capacités globales de l'organisme à éliminer le médicament. C'est un paramètre qui est propre à chaque individu car elle dépend de plusieurs facteurs physiologiques (comme la fonction rénale, la fonction hépatique, le poids, ou encore l'âge par exemple) mais aussi de facteurs génétiques (notamment la variabilité des enzymes hépatiques). En pratique, la pharmacocinétique de nombreux médicaments est linéaire et suit une cinétique d'ordre 1 c'est-à-dire qu'il y a une relation de proportionnalité entre la vitesse d'élimination et la quantité de médicament. Voici l'équation qui permet de calculer la clairance d'élimination totale dans le cas d'une pharmacocinétique linéaire qui suit une cinétique d'ordre 1 :

$$CL(t) = \frac{v_e(t)}{C(t)} = \frac{k \cdot V_d \cdot C(t)}{k_e \cdot C(t)} = k \cdot V_d$$

Le V_d correspond au volume de distribution qui sera abordé plus tard dans le manuscrit.

La clairance d'élimination totale résulte de plusieurs mécanismes d'élimination (Figure 19). Elle repose principalement sur l'élimination rénale et l'élimination hépatique, les deux principales voies d'élimination.

D'autres voies d'élimination, comme la voie pulmonaire ou encore l'élimination par d'autres tissus de l'organisme doivent également être prises en compte.

De manière rigoureuse, la clairance d'élimination totale peut se calculer avec cette formule :

$$CL = CL_R + CL_{NR}$$

Avec CL_R et CL_{NR} la clairance rénale et non rénale respectivement.

En pratique la CL_{NR} est représentée par la clairance hépatique car les deux voies d'élimination principales sont la voie rénale et la voie hépatique.

$$CL \approx CL_R + CL_H$$

Avec CL_H la clairance hépatique.

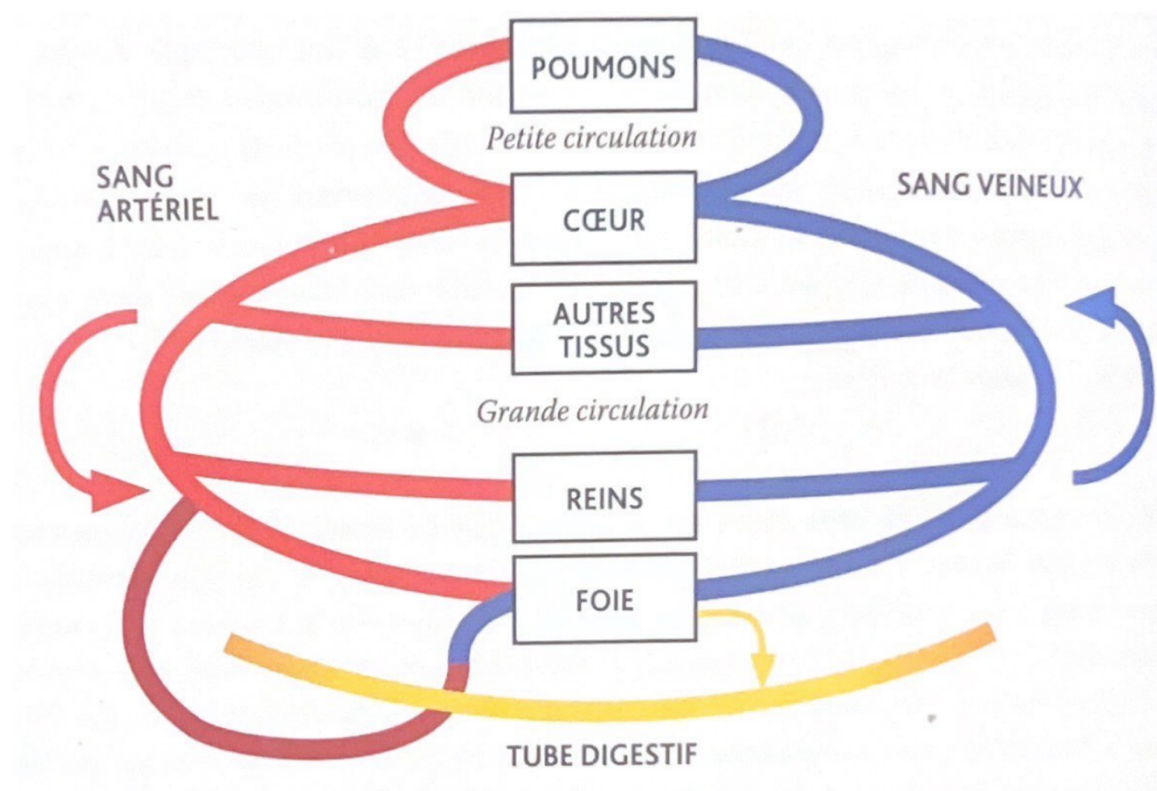


Figure 19 : Les différentes voies pouvant être impliquées dans l'élimination des médicaments³².

2.2.1.2.1.1. Clairance rénale

La clairance rénale se calcule façon similaire au calcul de la clairance plasmatique.

$$\frac{dQ_u}{dt} = v_{e,r}(t) = CL_R \cdot C(t)$$

La clairance rénale se calcule grâce à l'analyse des urines collectées après administration de la substance à étudier mais aussi grâce à l'analyse des concentrations plasmatiques de cette dernière. Dans ce cas, la clairance rénale peut être calculée de la façon suivante :

$$CL_R = \frac{U_\infty}{AUC}$$

Avec U_∞ qui est la quantité totale de médicament éliminé sous forme inchangée dans les urines.

La clairance rénale peut également être vue comme une fraction de la clairance totale.

$$CL_R = f_e \times CL$$

Avec f_e la fraction de la dose administrée et éliminée dans les urines sous forme inchangée.

2.2.1.2.1.2. Clairance métabolique hépatique

La clairance métabolique hépatique dépend de trois facteurs :

- Le débit sanguin hépatique qui va influencer la quantité de médicament qui atteint le foie par unité de temps (Q_H).
- La fraction libre plasmatique qui correspond au médicament non lié aux protéines plasmatiques (f_u)
- La clairance intrinsèque qui varie avec l'affinité de la substance pour les enzymes hépatiques impliquées dans la métabolisation de la substance étudiée (CL_{int}).

La clairance métabolique se calcule à l'aide de ces différents paramètres :

$$CL_H = \frac{f_u \cdot CL_{int} \cdot Q_H}{Q_H + f_u \cdot CL_{int}}$$

2.2.1.2.2. Le volume de distribution³²

Le volume de distribution (V_d) est un paramètre pharmacocinétique qui évalue la distribution tissulaire du principe actif (Figure 20). Il se calcule de la façon suivante :

$$V_d = \frac{Q(t)}{C(t)}$$

$Q(t)$ est la quantité de principe actif dans le plasma, $C(t)$ est la concentration plasmatique en fonction du temps. Ce paramètre est un volume qui est la plupart du temps proportionnel au poids de l'individu et peut donc être exprimé en L/kg. Le V_d peut se calculer directement à partir des concentrations plasmatiques.

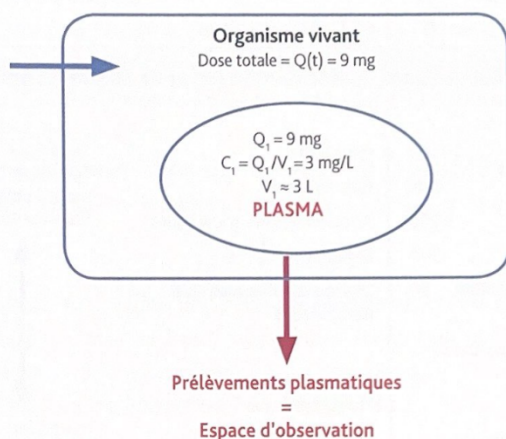
Si le médicament a été administré par voie intraveineuse (IV) alors le V_d sera calculé après l'administration. Tandis que, si le médicament n'a pas été administré par voie IV et que le coefficient de biodisponibilité F du

principe actif n'est pas connu alors ce n'est plus le V_d qui sera calculé mais le volume de distribution apparent.

Plus la concentration plasmatique est élevée moins le principe actif se distribue dans les tissus, ce qui veut donc dire que le volume de distribution est faible, et inversement. Le V_d permet de quantifier la distribution de la substance étudiée. Il représente le volume théorique qui contiendrait la quantité totale de principe actif présent dans l'organisme avec une concentration égale à la concentration plasmatique juste après l'injection, c'est donc un volume fictif.

Cas n° 1 : le principe actif ne se distribue pas dans les tissus : la quantité totale administrée, $Q(t)$, se répartit uniquement dans le volume V_1 .

$$V_d = \frac{Q(t)}{C_1} = \frac{9 \text{ mg}}{3 \text{ mg/L}} = 3 \text{ L}$$



Cas n° 2 : le principe actif se distribue à 99 % dans les tissus. La concentration mesurée dans le plasma est égale à C_1 .

La quantité totale administrée, $Q(t)$, se répartit dans le volume V appelé

$$V_d = \frac{Q(t)}{C_1} = \frac{9 \text{ mg}}{0,03 \text{ mg/L}} = 300 \text{ L}$$

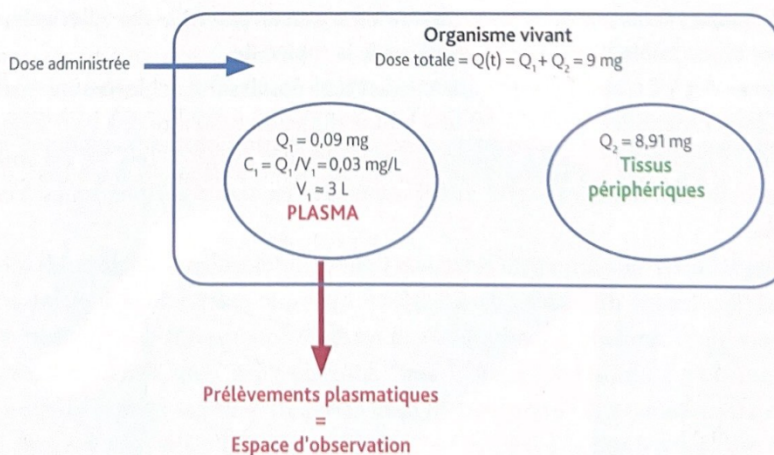


Figure 20 : Schéma illustrant la notion de volume de distribution³².

Enfin le V_d peut également se calculer à l'aide du volume plasmatique et du $V_{\text{eau hors plasma}}$ (volume d'eau total de l'organisme sans le plasma).

$$V_d = V_{plasma} + \frac{f_u}{f_{ut}} \times V_{eau\ hors\ plasma}$$

Avec f_{ut} la fraction libre de la substance étudiée dans les tissus et f_u la fraction libre dans le plasma celle qui n'est pas liée aux protéines plasmatiques. Il y a donc une relation entre liaison aux protéines plasmatiques et V_d . Plus la liaison aux protéines plasmatiques est importante plus le principe actif reste dans la circulation ce qui implique une diminution du V_d .

2.2.2. Modèle pharmacocinétique non linéaire et modèle populationnel

2.2.2.1. Modèle non linéaire³³

La pharmacocinétique non linéaire se caractérise par le fait qu'au moins un des paramètres fondamentaux varie en fonction de la concentration plasmatique en médicament. Cela se produit lorsqu'il y a des mécanismes saturables comme c'est le cas pour les nanomédicaments. Ces mécanismes peuvent être de différents types :

- Métabolisme saturable
- Sécrétion tubulaire rénale
- Fixation aux protéines plasmatiques

2.2.2.1.1. Métabolisme saturable

Le métabolisme saturable correspond au fait que certains processus métaboliques impliqués dans la métabolisation de la substance étudiée peuvent être saturés lorsque toutes les enzymes, transporteurs ou autres mécanismes intervenant dans la métabolisation de cette substance sont occupées. C'est le cas pour les substances ayant un métabolisme hépatique telles que les nanomédicaments.

La cinétique de ce type de métabolisme est définie par l'équation de Michaelis-Menten décrite de la façon suivante :

$$\frac{dQ(t)}{dt} = \frac{V_{max} \cdot C(t)}{K_m + C(t)}$$

La V_{max} est la vitesse maximale de métabolisation et dépend principalement du processus impliqué. Elle dépend également de la vitesse de réaction qui correspond à la quantité de substrat transformé par unité de temps. Le K_m est le reflet de l'affinité de la substance étudiée pour les processus impliqués dans la métabolisation de cette dernière. Un K_m faible indique une affinité élevée et inversement.

2.2.2.1.2. Sécrétion tubulaire rénale

Ce phénomène peut être présent lorsque la substance étudiée met en jeu des mécanismes de transport. Ce phénomène est possible si les concentrations de la substance ne sont pas très inférieures à la constante de dissociation (K_d). Cette constante de dissociation est le reflet de l'affinité de la substance pour le transporteur. Elle représente la concentration de la substance pour laquelle 50% des récepteurs sont occupés par des ligands.

2.2.2.1.3. Fixation aux protéines plasmatiques

La plupart du temps, ce mécanisme n'est pas saturable car ils présentent une constante de dissociation plutôt élevée qui implique un phénomène de saturation faible. Cependant, certains médicaments ont des concentrations très proches du K_d ce qui implique que l'affinité du médicament pour les protéines plasmatiques doit être prise en compte car un phénomène de saturation peut avoir lieu. La concentration plasmatique des médicaments liés aux protéines plasmatiques peut donc être calculée à l'aide de cette formule :

$$C_b = \frac{C_{b,max} \cdot C_u}{K_d + C_u}$$

C_u représente la concentration de médicaments sous forme libre tandis que $C_{b,max}$ correspond à la concentration maximale de médicaments pouvant être liés aux protéines plasmatiques.

Dans ce cas, f_u va donc augmenter avec les concentrations car les sites de liaison des protéines plasmatiques commencent à être de plus en plus occupés ce qui signifie que moins de médicament peut y être fixé. Il y a donc plus de médicament sous forme libre.

Or f_u influence la clairance de filtration glomérulaire et la clairance métabolique hépatique ainsi que le V_d . Une augmentation de f_u entrainera une augmentation des clairances ainsi qu'une augmentation du V_d .

2.2.2.2. Modèle populationnel^{34,50}

La pharmacocinétique de population (PKPOP) est une approche utilisée lorsqu'il n'y a pas beaucoup de données disponibles pour un patient donné (par exemple un nombre de prélèvements inférieur à 3).

Cette approche permet de compenser le manque d'informations en adoptant une approche populationnelle qui regroupe les informations de tous les individus ayant des données limitées. Pour obtenir un bon modèle, il faudrait avoir au moins une dizaine d'individus.

L'objectif principal de la PKPOP est de modéliser la relation entre un ensemble de variables indépendantes comme la dose, le temps, et certaines covariables (comme l'âge, le sexe etc) et des variables dépendantes à

savoir les concentrations des médicaments. Elle permet, à partir de cette analyse, de déterminer les paramètres moyens pour la population étudiée.

La PKPOP a deux types d'application principales :

- L'identification des covariables qui influencent les paramètres pharmacocinétiques (PK) de la substance étudiée ce qui permet de trouver des relations mathématiques entre les covariables et les paramètres PK.
- La détermination des paramètres PK d'un patient avec peu de données ne permettant pas une analyse individuelle classique. Ces paramètres sont obtenus par la méthode de l'estimation bayésienne.

La partie modélisation sera détaillée plus loin dans le manuscrit.

2.3. Intérêt de la technique FRET pour l'évaluation pharmacocinétique des NCL³⁵

La pharmacocinétique des nanomédicaments est encore peu étudiée car elle nécessite des techniques de marquage complexes qui préservent les qualités du nanomédicament tout en permettant de le suivre efficacement. Pour cela, Lebreton *et al.* ont réalisé une étude d'évaluation de la pharmacocinétique de différentes formulations de NCL grâce à une technique de dosage innovante la technique *Förster Resonance Energy Transfer* (FRET).

En effet, la technique FRET associée au comptage des NCL (Figure 21) a permis pour la première fois de doser les NCL sous forme intacte dans le sang. Le comptage des NCL est réalisé grâce à la technique d'analyse par suivi individuel des nanoparticules (Nanoparticle Tracking Analysis, NTA) permettant de déterminer une concentration initiale de la formulation. Celle-ci est directement reliée au signal FRET pour en déduire la concentration de particules intactes car cette méthode se base sur la mesure d'un ratio appelée *Proximity ratio* (PR) :

$$Proximity\ ratio(PR) = \frac{Longueur\ d'onde\ accepteur}{Longueur\ d'onde\ donneur + Longueur\ d'onde\ accepteur}$$

Pour utiliser cette technique il faut intégrer dans les NCL formulées deux agents fluorescents, ici le 1,1' -dioctadecyl-3,3,3' ,3' -tetramethylindocarbocyanine perchlorate et 1,10-dioctadecyl-3,3,30,30-tetramethylindocarbocyanine perchlorate (DiD et Dil respectivement) permettant de quantifier les NCL intactes dans les échantillons. Le Dil est capable d'entrer en état d'excitation lorsqu'il est soumis à une source énergétique. Tandis que le DiD est un agent capable de recevoir le signal énergétique du Dil excité si la distance qui les sépare est inférieure à 5 nm. Ainsi, l'intensité

du signal reçu du DiI permet d'évaluer l'intégrité des NCL à l'aide du calcul du PR. En effet, un PR proche de 1 indique que les NCL sont intactes tandis qu'un PR proche de 0 indique qu'elles sont cassées. L'intensité du signal permet d'évaluer la concentration de NCL intactes dans l'échantillon analysé.

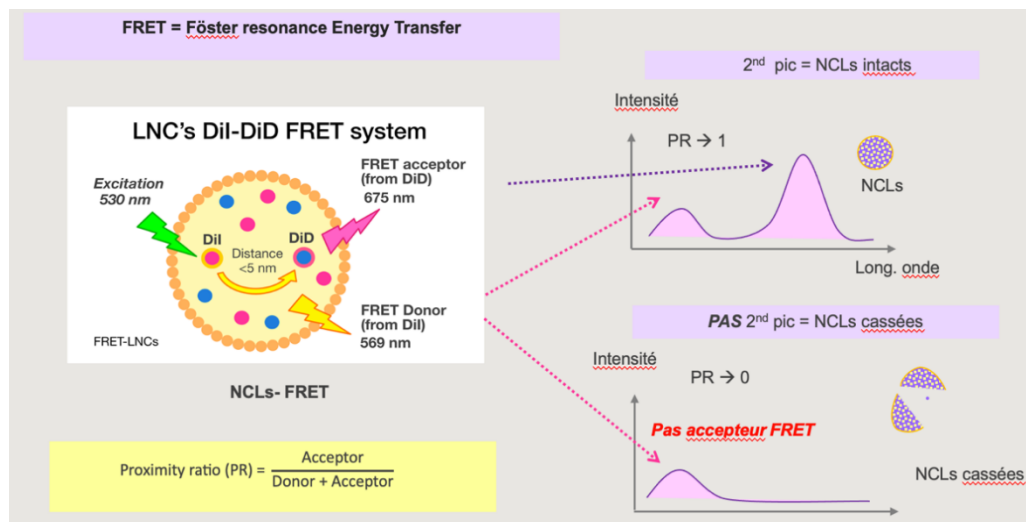


Figure 21 : Schéma illustrant la technique FRET appliquée d'après Roger *et al.* ³⁶.

6 formulations de différentes tailles et de différents types ont été réalisées à savoir :

- 6 NCL 50 normales
- 6 NCL 50 PEG
- 6 NCL 50 SA
- 6 NCL 85 normales
- 6 NCL 85 PEG
- 4 NCL 85 SA

Les compositions de chaque NCL sont données dans le tableau suivant (Tableau 1) :

Tableau 1 : Composition des NCL formulées³⁵.

Compositions	Quantity (% w/w)					
	LNC-50	LNC-50-PEG	LNC-50-SA	LNC-85	LNC-85-PEG	LNC-85-SA
Captex® 8000 (2% w/w DiI-TPB)	5.5	5.5	5.5	8.5	8.5	8.5
Captex® 8000 (2% w/w DiD-TPB)	5.5	5.5	5.5	8.5	8.5	8.5
Kolliphor® HS-15	11.5	11.5	11.5	9.3	9.3	9.3
Purified water	21.3	21.3	21.3	17.5	17.5	17.5
DSPE-mPEG-2000	–	0.6	–	–	0.6	–
Stearylamine	–	–	0.1	–	–	0.1
Lipoid® S75–3	0.7	0.7	0.7	0.7	0.7	0.7
NaCl	0.8	0.8	0.8	0.8	0.8	0.8
Purified water (2° C)	54.7	54.7	54.7	54.7	54.7	54.7

Mais également leur potentiel zêta qui représente la charge de la surface des NCL (Tableau 2).

Tableau 2 : Potentiel zêta des différentes NCL³⁵.

Formulations	Particle Size (nm)	Span	Zeta Potential (mV)	FRET Proximity Ratio (PR)	Particle Concentration (10 ² Tp/mL)
LNC-50 (a) (n = 5)	49.3 ± 4.8 ^d	0.51 ± 0.20	4.2 ± 2.1 ^{bcd}	0.90 ± 0.02	3.2 ± 0.8 ^b
LNC-50-PEG (b) (n = 5)	47.5 ± 2.3 ^e	0.40 ± 0.04 ^e	–7.1 ± 2.8 ^{ae}	0.87 ± 0.04	10.1 ± 1.1 ^{ae}
LNC-50-SA (c) (n = 7)	52.7 ± 4.9 ^f	0.56 ± 0.09	18.0 ± 1.0 ^{af}	0.86 ± 0.03 ^f	3.8 ± 0.5
LNC-85 (d) (n = 5)	84.8 ± 3.1 ^a	0.61 ± 0.04	16.9 ± 1.1 ^{ae}	0.90 ± 0.03	2.5 ± 0.8
LNC-85-PEG (e) (n = 3)	78.6 ± 2.1 ^b	0.67 ± 0.09 ^b	–1.3 ± 1.8 ^{bd}	0.93 ± 0.01	3.3 ± 0.5 ^b
LNC-85-SA (f) (n = 3)	83.1 ± 5.4 ^c	0.62 ± 0.03	30.0 ± 1.1 ^{cd}	0.93 ± 0.01 ^c	2.8 ± 0.8

Une analyse non compartimentale a été réalisée afin de calculer les paramètres pharmacocinétiques de chaque NCL formulées. Puis un modèle de population a été réalisé afin de décrire au mieux la pharmacocinétique des NCL ainsi que l'influence du type et de la taille sur cette dernière.

2.3.1. Analyse non compartimentale

Afin de calculer les paramètres pharmacocinétiques, les différentes NCL ont été injectées par voie IV chez 43 rats mâles de type Sprague Dawley dont 15 rats ont reçu une deuxième injection 7 jours après la première dose.

Puis les concentrations ont été mesurées grâce à la technique FRET à 10 temps de prélèvements qui sont : 5, 30, 60, 90, 120, 180, 240, 360, 600 et 1440 minutes. Certaines valeurs sont manquantes en raison de l'absence de prélèvement ou en raison des limites de quantification de la technique de dosage. Lorsque ces valeurs manquantes étaient terminales, elles ont été remplacées par la *Lower Limit Of Quantification* (LLOQ) qui est la limite de quantification de la technique de dosage (Tableau 3).

Tableau 3 : Tableau des LLOQ des différentes NCL³⁵.

	LNC-50	LNC-50-PEG	LNC-50-SA	LNC-85	LNC-85-PEG	LNC-85-SA
LLOQ (Tp/ mL) (m + sd)	16.7± 6.3	15.6± 6.8	13.1± 2.7	7.6± 2.3	16.6± 2.3	6.0± 2.4

De plus, les valeurs inférieures à cette LLOQ nommées *Below the Lower limit Of Quantification* (BLOQ) ont été supprimées comme le préconise la guideline « *guidance of analytical considération* » de la FDA.

L'analyse non compartimentale a permis de calculer les paramètres pharmacocinétiques (Tableau 4) :

Tableau 4 : Paramètres pharmacocinétiques des NCL³⁵.

PK parameters (NCA) 1st injection (D0)/2nd injection (D7)	LNCs 50 nm			LNCs 85 nm			
	LNC-50 (a) (n = 3)/(n = 3)	LNC-50-PEG (b) (n = 3)/(n = 3)	LNC-50-SA (c) (n = 3)/(n = 3)	LNC-85 (d) (n = 3)/(n = 3)	LNC-85-PEG (e) (n = 3)/(n = 3)	LNC-85-SA (f) (n = 3)/(n = 1)	
AUC _{0-∞} (T _p • min/mL)	D0	22016 ^b [18,172, 25,860]	9432 ^{ac} [4524, 14,340]	17,733 [10,282, 25,184]	15984 ^e [8767, 23,200]	27672 ^{bd} [21,890, 33,454]	17,091 [9369, 24,813]
	D7	14,376 [0, 30,261]	7906 [0, 16,814]	13,569 [4125, 23,012]	18,318 [9688, 26,948]	26,858 [18,578, 35,138]	14,217
AUC _{0-last} (T _p • min/mL)	D0	12768 ^b [7622, 17,914]	6871 ^{ac} [6034, 7708]	11,441 [9182, 13,699]	13505 ^e [9664, 17,346]	18744 ^{bd} [17,982, 19,506]	15,198 [8996, 21,400]
	D7	8623 [4220, 13,027]	5010 [3493, 6528]	7968 [4357, 11,580]	15,563 [9565, 21,560]	19,927 [16,419, 23,425]	12,201
Half-life (min)	D0	312 ^b [25, 599]	102 ^{ac} [5, 199]	252 [53, 451]	240 [86, 394]	358 ^b [243, 474]	202 [41, 362]
	D7	441 ^b [0, 610]	143 ^{ac} [0, 239]	353 [21, 514]	323 [129, 313]	522 ^b [212, 375]	272 [72, 471]
MRT _{obs} (min)	D0	24.8 [28, 854]	16.8 [3, 283]	26.6 [84, 623]	17.2 [128, 519]	13.9 [347, 697]	22.2 [72, 471]
	D7	22.5 [0, 871]	20.1 [0, 378]	26.6 [51, 686]	17.2 [196, 428]	13.9 [308, 551]	22.2 [300, 0]
Cl _{obs} (10 ⁻³ mL/min)	D0	5.5 ^b [4.5, 6.2]	12.1 ^{ac} [5.5, 18.7]	6.3 [3.1, 9.4]	5.6 [2.8, 8.3]	3.0 ^b [2.2, 3.8]	5.8 [3.6, 7.9]
	D7	7.4 [0, 16.8]	14.0 [1.4, 26.6]	7.3 [2.4, 12.3]	5.5 [2.5, 8.5]	3.3 [1.9, 4.7]	7.1 [1.9, 4.7]
V _D (mL)	D0	24.8 [0.7, 48.8]	16.8 [9.6, 23.9]	26.6 [11.8, 32.3]	17.2 [14.8, 22.5]	13.9 [11.7, 19.2]	22.2 [5.1, 27.7]
	D7	22.5 [18.4, 26.5]	20.1 [18.8, 21.4]	26.6 [16.4, 36.8]	17.2 [12.4, 21.9]	13.9 [11.6, 16.1]	22.2

T_p = 10¹² particles (Teraparticle). Statistical analysis was conducted by two-way ANOVA and is expressed if the level of significance ≤ p 0.05 as: (a) vs LNC-50, (b) vs LNC-50-PEG, (c) vs LNC-50-SA, (d) vs LNC-85, (e) vs LNC-85-PEG, (f) vs LNC-85-SA. A statistical analysis was conducted by paired t-test to compare the first injection and second injection.

Les courbes de l'évolution des concentrations en fonction du temps ont également été calculées (Figure 22) :

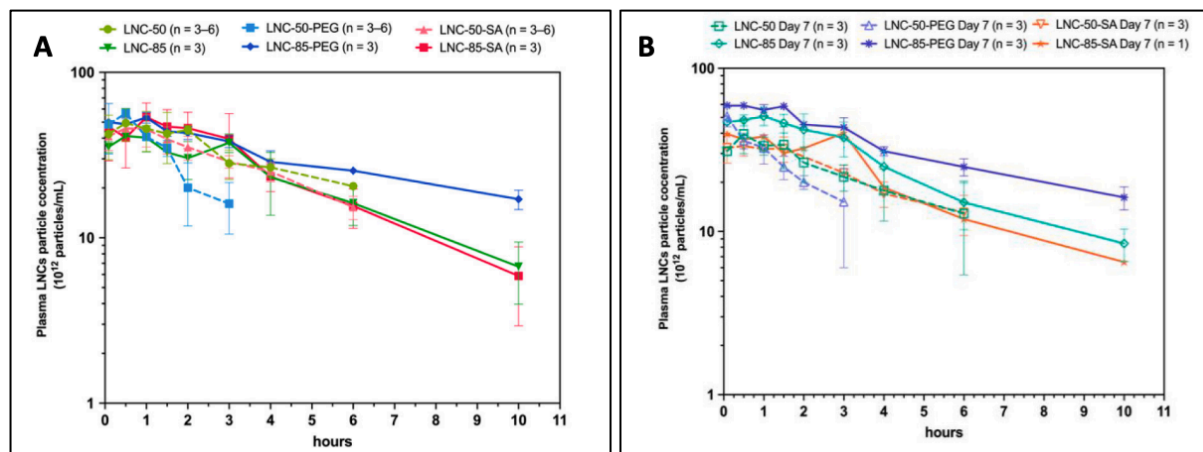


Figure 22 : Évolution des concentrations plasmatiques des différentes NCL en fonction du temps après injection à J0 (A) et à J7 (B)³⁵.

Le graphique A représente l'évolution des concentrations plasmatiques des différentes NCL en fonction du temps après l'injection J0. Le graphique B représente l'évolution des concentrations plasmatiques des différentes NCL en fonction du temps avec l'injection 7 jours après la première injection.

Cette étude a montré que les NCL 85-PEG offrent l'exposition la plus élevée tandis que les NCL 50-PEG ont l'exposition la plus faible. Concernant l'élimination, les NCL qui s'éliminent le plus vite sont les NCL 50-PEG et celles qui s'éliminent le plus lentement sont les NCL 85-PEG. Les NCL ayant le temps de séjour moyen le plus long sont les NCL 85 PEG et celles ayant le temps de séjour le plus court sont les NCL 50 PEG. Enfin, les NCL qui se distribuent le mieux dans les tissus sont les NCL 50 Norm et celles qui se distribuent le moins sont les NCL 85 Norm.

2.3.2. Modèle de population

Un modèle de population a été développé avec les individus ayant un nombre de prélèvements supérieur ou égal à 5 (Tableau 5). Ce modèle populationnel a été développé avec une bonne précision (erreur standard relative < 20%) et a permis d'évaluer l'impact de la taille et du type de la NCL sur la pharmacocinétique.

Au total, le modèle final est de type mono-compartmental avec une cinétique d'élimination de type Michaelis-Menten. Les concentrations sont estimées avec une erreur qui est considérée comme proportionnelle

Ces notions d'erreur et de compartiments seront décrites plus tard dans ce travail.

Ce modèle a d'ailleurs démontré que l'élimination dépend de la taille et du type des NCL alors que la distribution dépend uniquement de la taille de ces derniers. En effet, il a été démontré que le V_d des NCL de 85 nm est inférieur à celui des NCL de 50 nm ; que les NCL 85-PEG ont une clairance inférieure à celles des autres NCL et que les NCL 50 PEG ont une élimination plus rapide que les autres formulations en raison d'une instabilité sanguine.

Tableau 5 : Modèle de population décrivant la pharmacocinétique des NCL chez les rats de type Sprague Dawley³⁵.

Parameters	BASE MODEL		FINAL MODEL		
	Estimate	RSE (%)	Estimate	RSE (%)	Typical value
<i>Fixed Effects</i>					
V_D (mL)	20.29	3.30	21.49	3.68	21.49
$\beta_{85_V_D}$	NA	NA	-0.16	36.00	18.31
V_m (Tp/min)	3.93	8.89	3.89	4.19	3.89
$\beta_{LNC-85_V_m}$	NA	NA	-0.30	25.60	2.88
$\beta_{LNC-50-PEG_V_m}$	NA	NA	0.75	9.21	8.24
$\beta_{LNC-85-PEG_V_m}$	NA	NA	-0.74	12.10	1.86
$\beta_{LNC-85-SA_V_m}$	NA	NA	-0.20	41.00	3.18
K_m (Tp/mL)	27.00	NA	27.00	NA	27.00
<i>Standard Deviation of the Random Effects</i>					
ω_{V_D}	0.15	19.80	0.13	17.50	NA
ω_{V_m}	0.46	18.00	0.11	13.70	NA
<i>Error Model Parameters</i>					
b	0.20	4.73	0.20	4.68	NA

3. Limited Sampling Strategies (LSS)

3.1. Principe général

La *limited sampling strategies* est une méthode d'échantillonnage limité qui a pour but de réduire le nombre d'échantillons nécessaires pour prédire l'exposition d'un médicament (AUC). Cette réduction du nombre d'échantillon ne doit pas affecter la précision et le biais du modèle, ce qui implique que les échantillons sélectionnés doivent être suffisamment représentatifs de la population étudiée. Cette technique est principalement mise en place lorsque certains échantillons sont difficiles et/ou longs à récolter ou lorsque le coût de l'analyse des échantillons est onéreux par exemple³⁷.

À titre d'exemple supposons un médicament anticancéreux en phase d'étude clinique. Afin de mettre ce médicament sur le marché, il est

nécessaire d'étudier son profil pharmacocinétique chez les patients. Cette étude permettra de contrôler l'exposition du patient au médicament étudié mais également de définir des protocoles de suivi des concentrations plasmatiques vérifiant que ces dernières sont comprises dans la marge thérapeutique. Pour établir le profil pharmacocinétique de cet anticancéreux, 10 prélèvements seront réalisés par exemple, chez les patients. Une stratégie de LSS permettrait par exemple de récolter ces 10 prélèvements auprès d'un échantillon restreint mais représentatif de la population de patients. À partir de ces prélèvements collectés, la LSS pourrait par exemple limiter le nombre de prélèvements à 3 par sujet, en s'appuyant sur un modèle mathématique permettant d'estimer l'exposition grâce à ces seuls prélèvements (Figure 23). Le modèle ainsi obtenu pourra être utilisé pour d'autres patients s'il a démontré de bonnes performances.

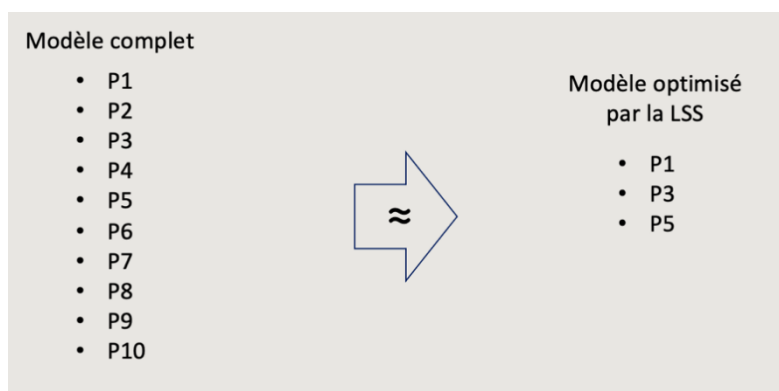


Figure 23 : Exemple de LSS à 3 prélèvements.

Pour réaliser la LSS, de nombreuses étapes sont nécessaires. Tout d'abord, les données doivent être analysées et traitées pour que les modèles puissent être performants. Puis intervient la sélection des variables qui permet de choisir les variables les plus contributives à la prédiction de la variable cible. S'en suit une étape de modélisation avec ces variables sélectionnées puis la validation et le choix du modèle le plus performant grâce à des techniques de validation spécifiques. Une fois ces étapes réalisées, le modèle peut être appliqué en clinique.

3.2. Traitement des données

3.2.1. Exploration des données

3.2.1.1. Statistiques exploratoires

3.2.1.1.1. Effectif

L'effectif est un paramètre important qui désigne le nombre d'individus d'une base de données. Il se calcule en comptant le nombre d'individus dans la base de données.

L'effectif permet de calculer de nombreux indicateurs de performance du modèle comme le r^2 , la *root mean squared error* (précision du modèle), la *mean predictive error* (biais du modèle), etc. Mais également des

paramètres d'exploration des données comme la moyenne, la médiane, l'écart type, la variance, etc (exemple détaillé dans le Tableau 6).

De plus, la quantité de données est importante car elle permet également d'évaluer la robustesse du modèle. En effet, plus il y a de données, plus le modèle aura de chance de pouvoir comprendre au maximum le phénomène étudié, et inversement, moins le modèle aura de données plus il lui sera difficile de généraliser et d'effectuer de bonnes prédictions.

Tableau 6 : Exemple d'une base de données d'effectif n=10

Individus	Revenus annuel (en euros)
1	20 000
2	20 000
3	20 000
4	20 000
5	20 000
6	20 000
7	20 000
8	20 000
9	20 000
10	1 000 000

Dans cet exemple, l'effectif noté N est égal à 10 car il y a 10 individus dans la base de données.

3.2.1.1.2. Moyenne arithmétique³⁸

La moyenne arithmétique est un indicateur quasi-systématiquement calculé en statistique exploratoire et représente ce qu'on appelle la tendance centrale.

La moyenne arithmétique notée \bar{x} se calcule de cette façon :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n p_i x_i$$

Avec p_i les poids associés aux valeurs x_i tel que $\sum_{i=1}^n p_i = 1$.

Cependant, l'inconvénient de la moyenne est qu'elle est très sensible aux valeurs extrêmes. Par exemple si l'objectif est d'étudier le niveau de revenus de la population, et que sur les 10 personnes interrogées 9 gagnent 20000 euros par an et 1 gagne 1 million d'euros par an. Le calcul de la moyenne de cet échantillon prélevé dans la population donne :

$$\bar{x} = \frac{20\,000 \text{ d'euros} \times 9 + 1\,000\,000 \text{ d'euros}}{10} = 118\,000 \text{ euros}$$

Ce qui voudrait dire qu'en moyenne les individus de cet échantillon gagnent 118 000 euros par an ce qui n'est pas représentatif des 9 personnes qui gagnent 20 000 euros par an. Les valeurs extrêmes impactent donc fortement le calcul de la moyenne.

Il existe d'autres méthodes de calcul de la moyenne telles que la moyenne géométrique ou encore la moyenne harmonique.

3.2.1.1.3. Médiane³⁸

La médiane est également un indicateur qui représente la tendance centrale. Elle a pour caractéristique d'être moins sensible aux valeurs extrêmes que la moyenne car elle sépare les données en deux groupes dont 50% des valeurs sont en dessous de la médiane et 50% des valeurs sont au-dessus.

La médiane se calcule sur les données triées par ordre croissant et dépend de l'effectif. Si l'effectif est impair alors la médiane notée M se calcule de la façon suivante :

$$M = x_{\left(\frac{n+1}{2}\right)}$$

Dans le cas où l'effectif est pair, la médiane se calcule de la façon suivante :

$$M = \frac{x_{n/2} + x_{\frac{n}{2}+1}}{2}$$

En reprenant le même jeu de données, un tri des observations dans l'ordre croissant du Tableau 6 doit être effectué. Ce tri permet de calculer la médiane pour n = 10 qui est pair :

$$M = \frac{x_{10/2} + x_{\frac{10}{2}+1}}{2} = \frac{x_5 + x_6}{2} = \frac{20\,000 \text{ euros} + 20\,000 \text{ euros}}{2} = 20\,000 \text{ euros}$$

La médiane donne une valeur plus représentative que la moyenne et représente souvent mieux l'échantillon.

3.2.1.1.4. Valeur minimale et maximale

La valeur minimale et maximale d'une variable se définissent respectivement comme la plus petite valeur de cette variable et la plus grande valeur de cette variable. Ce sont des valeurs importantes notamment pour la détection des valeurs extrêmes ou encore pour avoir une première photographie du jeu de données.

Dans le Tableau 6, la valeur minimale de la variable du revenu annuel est 20 000 euros et la valeur maximale est 1 000 000 euros.

3.2.1.1.5. Variance^{38,39}

La variance notée s^2 est un indicateur de dispersion, elle représente l'écart moyen par rapport à la moyenne d'une variable. Son unité est l'unité de la variable au carré. Elle se calcule de la façon suivante :

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n p_i (x_i - \bar{x})^2$$

En reprenant l'exemple du Tableau 6, la variance du jeu de donnée est égale à :

$$s^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x})^2$$

$$s^2 = \frac{1}{10} \times (9 \times (20\,000 - 118\,000)^2) + (1\,000\,000 - 118\,000)^2$$

$$s^2 = 8.6436 \times 10^{10} \text{ euros}^2$$

3.2.1.1.6. Écart type^{39,38}

L'écart type noté σ est également un paramètre de dispersion. À la différence de la variance, il a la même unité que la variable étudiée ce qui facilite son utilisation et son interprétation. Il est égal à la racine carrée de la variance.

$$\sigma = \sqrt{s^2} = s$$

En reprenant l'exemple du Tableau 6, l'écart type serait égal à :

$$\sigma = \sqrt{8.6436 \times 10^{10}} = 29\,400 \text{ euros}$$

3.2.1.1.7. Quartiles ³⁹

Les quartiles sont des indicateurs permettant de séparer les données en plusieurs parties. Le premier quartile est calculé de telle sorte que l'on retrouve 75% des observations en dessous d'une certaine valeur et 25% des valeurs au-dessus de cette valeur. Tandis que le troisième quartile est la séparation des données telles que 25% des valeurs sont en dessous d'une certaine valeur et 75% des données au-dessus de cette valeur. La connaissance de ces quartiles permet d'avoir une meilleure appréciation de la répartition des données.

La connaissance de ces quartiles est importante car elle permet de calculer l'étendue interquartile (EIQ) qui est également une mesure de dispersion³⁸ dont la formule est :

$$EIQ = Q_3 - Q_1$$

Dans notre exemple, $Q_1 = 20\,000$ euros car 75% des valeurs sont inférieures ou égales à 20 000 euros et 25% sont supérieures ou égales à 20 000 euros. $Q_3 = 20\,000$ euros car 25% des valeurs sont inférieures ou égales à 20 000 euros et 75 % sont supérieures ou égales à 20 000 euros. L'EIQ peut donc se calculer de la façon suivante :

$$EIQ = 20\,000 - 20\,000 = 0 \text{ euros}$$

3.2.1.2. Visualisation des données

3.2.1.2.1. Nuage de points

Le nuage de points consiste à représenter une variable par rapport à une autre sur un graphique en deux dimensions. Ce graphique permet de voir les relations qu'il y a entre deux variables. La visualisation permet également de détecter les relations qu'il y a entre les variables étudiées. Un exemple de nuage de points est représenté dans la Figure 24 ci-dessous.

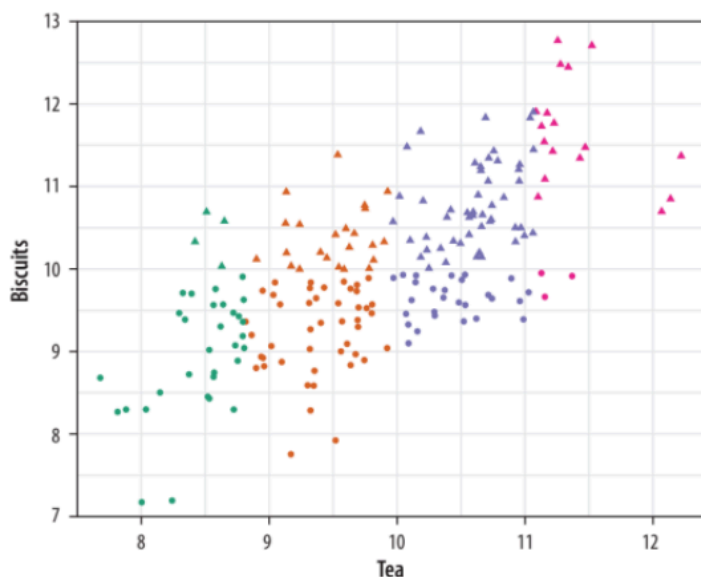


Figure 24 : Exemple de nuages de points qui représente la variable Biscuits en fonction de la variable Tea⁴⁰.

3.2.1.2.2. Histogrammes³⁸

Un histogramme est un graphique sous forme de barres verticales représentant les différentes classes d'observations. La surface de ces barres est proportionnelle à la fréquence de la classe (exemple de la Figure 25).

Il n'y a pas de règles définies pour établir le découpage de classe. Cependant, un trop grand nombre de classe ou un trop faible nombre de classe aboutit à des graphiques incohérents.

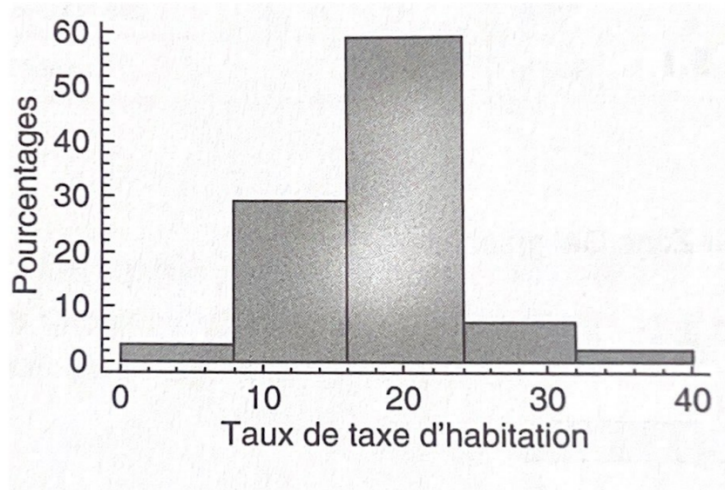


Figure 25 : Exemple d'un histogramme d'une variable représentant le taux de taxe d'habitation³⁸.

3.2.1.2.3. Box plots³⁸

Les box plots appelés aussi boîte à moustaches sont une représentation synthétique des caractéristiques d'une variable numérique (exemple de la Figure 26).

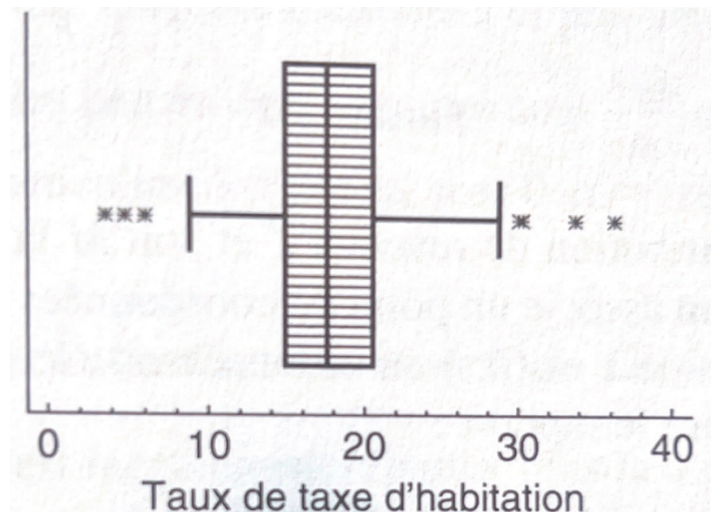


Figure 26 : Exemple d'une boîte à moustaches de la variable taux d'habitation³⁸.

La boîte à moustaches est la partie centrale de la distribution elle comprend la moitié des valeurs comprises entre le premier quartile et le troisième quartile.

Les moustaches s'étendent des deux côtés de cette boîte à gauche de $Q_1 - 1,5 \times (Q_3 - Q_1)$ et à droite de $Q_1 + 1,5 \times (Q_3 - Q_1)$. Les valeurs au-delà de ces moustaches représentent des valeurs hors normes, suspectes ou aberrantes. Les valeurs aberrantes si présentes peuvent faire l'objet de nombreux traitements. Elles peuvent soit être supprimées du jeu de données soit transformées pour réduire leur impact.

3.2.2. Traitement des valeurs manquantes⁴¹

Les valeurs manquantes sont des valeurs qui ne sont pas référencées dans la base de données alors qu'elles auraient dû l'être. Plusieurs causes peuvent expliquer la présence de valeurs manquantes :

- Oubli de récolte de la donnée
- Pas de valeurs à référencer
- Donnée erronée

Le traitement des valeurs manquantes est essentiel pour plusieurs modèles notamment la régression linéaire multiple qui ne peut pas travailler avec des valeurs manquantes.

Les méthodes d'imputation les plus connues sont l'imputation par la moyenne, l'imputation par la médiane et l'imputation par une constante. Il existe cependant de nombreuses méthodes d'imputation comme l'imputation à l'aide de modèles ou l'imputation à l'aide d'équations personnalisées.

3.2.2.1. Imputation par une constante

La méthode de l'imputation de la constante consiste à remplacer les valeurs manquantes par une constante censée représenter au mieux la valeur manquante.

L'inconvénient de cette méthode est qu'il faut choisir une constante qui conviendra le mieux, ce qui revient à faire la meilleure approximation de la valeur manquante avec les connaissances des experts par exemple.

Lorsque les connaissances métiers ne permettent pas de faire cela, il faudrait tester plusieurs constantes ce qui peut être chronophage et coûteux en termes de calcul informatique.

Par exemple, supposons que 5 personnes aient été interrogées sur le nombre de médicaments qu'ils doivent prendre par jour mais que la personne numéro 3 n'ait pas voulu répondre au questionnaire (valeur manquante). Le tableau suivant recense les différents résultats du sondage (Tableau 7).

Tableau 7 : Exemple des résultats d'un sondage avec valeur manquante auprès de 6 personnes concernant le nombre de médicaments qu'ils prennent par jour.

Individus	Nombre de médicaments à prendre par jour
1	5
2	3
3	?
4	8
5	7
6	5

L'imputation par la méthode de la constante pourrait se faire judicieusement par une constante entière car le nombre de médicament est un entier. De plus, il est judicieux d'imputer la valeur par une constante comprise dans l'intervalle entre la valeur minimale et la valeur maximale. Dans cet exemple (Tableau 8) l'imputation pourrait être une constante entière dans l'intervalle [3 ; 8] ce qui laisse le choix entre les chiffres : 3, 4, 5, 6, 7, 8. En modélisation, toutes ces imputations peuvent être testées pour trouver la constante qui donne les meilleurs résultats.

Tableau 8 : Exemple des résultats d'un sondage auprès de 6 personnes concernant le nombre de médicaments qu'ils prennent par jour imputée par la valeur 5.

Individus	Nombre de médicaments à prendre par jour
1	5
2	3
3	5
4	8
5	7
6	5

3.2.2.2. Imputation par la moyenne des valeurs

La méthode de l'imputation par la moyenne consiste à remplacer la valeur manquante par la moyenne des valeurs de la variable. En reprenant l'exemple précédent, la moyenne de la variable nombre de médicament à prendre par jour est égale à 5.6 médicaments par jour. Le tableau de données après imputation par la moyenne est le suivant (Tableau 9) :

Tableau 9 : Exemple des résultats d'un sondage auprès de 6 personnes concernant le nombre de médicaments qu'ils prennent par jour imputée par la méthode de la moyenne.

Individus	Nombre de médicaments à prendre par jour
1	5
2	3
3	5,6
4	8
5	7
6	5

3.2.2.3. Imputation par la médiane des valeurs

La méthode de l'imputation par la médiane consiste à remplacer la valeur manquante par la médiane des valeurs de la variable. En reprenant l'exemple précédent, la médiane de la variable nombre de médicament à prendre par jour est égale à 5 médicaments par jour. Le tableau de données après imputation par la médiane est le suivant (Tableau 10) :

Tableau 10 : Exemple des résultats d'un sondage auprès de 6 personnes concernant le nombre de médicaments qu'ils prennent par jour imputée par la méthode de la médiane.

Individus	Nombre de médicaments à prendre par jour
1	5
2	3
3	5
4	8
5	7
6	5

3.2.3. Mise à l'échelle

La mise à l'échelle est une étape importante car elle permet de mettre à l'échelle des variables ayant des dimensions différentes. Par exemple le prix d'une maison en euros et l'ancienneté de la maison en année n'ont pas les mêmes dimensions ni les mêmes plages de valeurs.

En effet, certains modèles ont besoin d'avoir des variables ayant les mêmes dimensions comme c'est le cas de la régression linéaire multiple.

La mise à l'échelle permet aux algorithmes de converger plus rapidement car il est plus facile pour l'algorithme de trouver la meilleure solution au problème d'optimisation. Elle permet aussi la comparaison des coefficients des variables entre eux car la mise à l'échelle permet de faire en sorte que chaque variable contribue équitablement au calcul des coefficients du modèle.

Les deux méthodes principales de mise à l'échelle sont la standardisation et la normalisation.

3.2.3.1. Standardisation⁴²

La standardisation est une méthode de mise à l'échelle qui permet de rendre la variable normalement distribuée avec une moyenne de 0 et un écart type de 1. Cette mise à l'échelle a pour objectif de rendre les variables centrées réduites comme la loi normale centrée réduite. C'est une méthode qui fonctionne mieux lorsque le modèle a besoin de données qui suivent une loi normale. De plus, elle a l'avantage d'être peu affectée par les valeurs aberrantes.

Il existe de nombreuses méthodes de standardisation, l'une des plus fréquentes utilise la formule :

$$x' = \frac{x - \mu}{\sigma}$$

Par exemple, ci-dessous un jeu de données représenté dans le Tableau 11 composé du poids et de l'âge des individus :

Tableau 11 : Exemple d'un jeu de données comportant les variables poids et âge.

Individus	Poids (kg)	Age (années)
1	56	15
2	73	20
3	87	23
4	102	42
5	82	46
6	75	35

Le poids et l'âge ne sont pas à la même échelle, le poids va de 56 à 102 kgs tandis que l'âge va de 15 à 46 ans. Les plages de valeurs ne sont pas les mêmes. Après standardisation, et application de la formule, les valeurs standardisées permettent d'obtenir un nouveau tableau (Tableau 12) :

Tableau 12 : Jeu de données comportant les variables poids et âges standardisés.

Individus	Poids (kg)	Age (années)
1	-1,64998	-1,31443
2	-0,4392	-0,8811
3	0,557905	-0,6211
4	1,626234	1,025545
5	0,201796	1,372208
6	-0,29676	0,418884

3.2.3.2. Normalisation⁴³

Cette méthode permet d'encadrer les variables entre 0 et 1. Plusieurs méthodes de normalisation existent mais la plus connue est la normalisation min – max. Pour cela la formule est utilisée :

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Cette méthode est sensible aux valeurs aberrantes car leur présence peut modifier radicalement la valeur des données. Elle est principalement utilisée lorsqu'il est souhaité que les valeurs soient comprises entre 0 et 1.

En reprenant les mêmes données que précédemment un nouveau tableau après application de la normalisation (Tableau 13) :

Tableau 13 : Jeu de données comportant les variables poids et âge normalisés.

Individus	Poids (kg)	Age (années)
1	0	0
2	0,369565	0,16129
3	0,673913	0,258065
4	1	0,870968
5	0,565217	1
6	0,413043	0,645161

3.2.4. Transformation de variables

La transformation des variables est une technique qui permet de modifier la distribution d'une variable afin qu'elle donne plus d'informations pour la construction du modèle. En effet, cette technique applique une fonction mathématique sur la variable afin que le modèle puisse capter plus d'information de cette dernière.

La transformation de variable s'utilise principalement lorsque les données ne suivent pas une loi normale afin de les rendre normalement distribuées. Les méthodes de transformation de variables les plus connues sont la transformation Box-Cox, la transformation puissance et la transformation logarithmique.

3.2.4.1. Transformation Box-Cox^{44,45}

La transformation de Box-Cox est une méthode qui applique une transformation selon les équations :

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(y) & \text{si } \lambda = 0 \end{cases}$$

Avec y la variable à transformer et le paramètre lambda (λ) à optimiser.

Cette transformation est utilisée pour avoir une distribution normale. Cependant un des inconvénients de cette transformation est qu'elle est valable que pour les données positives. Lorsque les valeurs sont négatives il faut plutôt utiliser une transformation de Yeo-Johnson qui est semblable à quelques paramètres près à la transformation de Box-Cox.

3.2.4.2. Transformation puissance

La transformation puissance est une méthode qui consiste à appliquer une fonction puissance à la variable à transformer. La nouvelle variable notée y' devient :

$$y' = y^\alpha$$

Avec α qui est le paramètre représentant la puissance appliquée. Si $\alpha = 2$, la transformation puissance est une transformation au carré. L'inconvénient de cette méthode réside dans la difficulté de la détermination du bon paramètre α .

3.2.4.3. Transformation logarithmique

La transformation logarithmique consiste à appliquer la fonction logarithme aux données. La nouvelle variable notée y' devient :

$$y' = \ln(y)$$

3.2.5. Analyse en composante principale

L'analyse en composante principale (ACP) est une méthode de réduction de la dimensionnalité des variables. Cette méthode permet de réduire le nombre de variables tout en réduisant la colinéarité entre les variables. En effet, la colinéarité peut être néfaste pour les modèles car elle entraîne des calculs biaisés et des résultats instables.

L'ACP vise à passer d'un ensemble de données à p variables à un nouvel ensemble de données à p' variables telles que les nouvelles variables créées

soient des combinaisons linéaires des variables initiales et qu'idéalement $p' < p$.

Les nouvelles variables créées sont alors appelées composantes principales et sont calculées de la façon suivante :

$$\textit{Composante principale } i = X_{\text{centré réduit}} \cdot v_i$$

Avec $X_{\text{centré réduit}}$ les variables explicatives centrées réduites et v_i le vecteur propre de la matrice $X_{\text{centré réduit}}$.

De cette relation, il est aisé de déduire cette formule qui montre que les composantes principales sont des combinaisons linéaires des variables explicatives.

$$\textit{Composante principale } i = \sum_{i=1}^n \beta_i X_i$$

Avec β_i les coefficients du vecteur propre v_i et X_i les variables initiales⁴⁶.

3.2.6. Sélection de variables

Dans le cadre de la LSS, la sélection de variables est très importante. En effet, le but de la LSS est de réduire le nombre de prélèvements donc le nombre de variables.

La sélection de variables permet de choisir les variables les plus pertinentes pour notre modèle et de supprimer celles qui ne le sont pas. Plusieurs approches existent : la sélection par le coefficient de corrélation, la méthode exhaustive qui permet de tester toutes les combinaisons possibles et les méthodes pas à pas qui consistent à sélectionner les variables selon l'importance de ces dernières.

3.2.6.1. Coefficient de corrélation⁴⁷

Les coefficients de corrélation servent à étudier les liaisons entre les variables. Lorsqu'une variable semble suivre une relation linéaire en fonction d'une autre, il est judicieux de calculer le coefficient de Bravais-Pearson tandis que si la relation est non-linéaire, il est préférable de calculer le coefficient de Spearman.

Des variables fortement corrélées entre elles nécessitent la suppression de l'une des deux variables car cela veut dire qu'elles apportent la même information ceci pouvant nuire aux performances du modèle construit.

3.2.6.1.1. Coefficient de Bravais-Pearson

Le coefficient de Bravais-Pearson encore appelé coefficient de corrélation linéaire noté r est un coefficient qui mesure le caractère linéaire entre deux variables. Il se calcule de la façon suivante :

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

C'est un coefficient compris entre -1 et 1. Un coefficient proche de 1 ou de -1 indique une très forte corrélation entre les variables. Si le coefficient est égal à 1 cela veut dire qu'il y a une très forte corrélation linéaire entre les deux variables et que l'augmentation d'une variable entraîne l'augmentation de l'autre variable. Si le coefficient est égal à -1 cela veut aussi dire qu'il y a une très forte corrélation entre les deux variables mais que l'augmentation de l'une des variables entraîne la diminution de l'autre. Un coefficient de corrélation nul indique qu'il n'y aurait pas de relation linéaire entre les deux variables étudiées.

3.2.6.1.2. Coefficient de Spearman

Le coefficient de Spearman s'interprète de la même façon que le coefficient de corrélation linéaire cependant il est plutôt utilisé pour des relations non-linéaires. Pour le calculer, il faut attribuer un rang de 1 à n aux deux variables explicatives étudiées (1 étant le rang de la valeur la plus haute de cette variable).

Notons d_i la différence entre les rangs des deux variables explicatives étudiées d'une même observation. Le coefficient de corrélation de Spearman se calcule alors à l'aide de cette formule :

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

3.2.6.2. Méthode exhaustive⁴⁸

La méthode exhaustive permet de tester toutes les combinaisons possibles de variables. Pour avoir toutes les possibilités de combinaisons de variables il faut calculer le nombre de combinaison de 1 variable parmi n variables ; de 2 variables parmi n variables, et ce jusqu'à n variables parmi n variables.

Pour calculer le nombre de combinaisons de k variables parmi n variables, la formule utilisée est la formule du coefficient binomial et se note de la façon suivante :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Pour obtenir toutes les combinaisons possibles, il faut faire la somme des coefficients binomiaux pour k allant de 1 jusqu'à n ce qui donne la formule suivante :

$$C = \sum_{k=1}^n \binom{n}{k} = \sum_{k=1}^n \frac{n!}{k!(n-k)!} = 2^n - 1$$

Si n est très grand, le nombre de combinaison est exponentiel ce qui augmente le temps de calcul et prend de la place en termes de stockage. Ce qui explique que la méthode exhaustive est principalement utilisée lorsque n est faible.

3.2.6.3. Méthodes pas à pas⁴⁸

Les méthodes pas à pas sont utilisées lorsque le nombre de variables est très élevé. Ce sont des méthodes qui utilisent des approches statistiques pour sélectionner les variables en construisant des modèles à l'aide des variables les plus informatives et suppriment celles qui le sont moins.

L'inconvénient principal de ces méthodes réside dans la sélection des variables qui peuvent être considérées comme non informatives alors qu'elles le sont en réalité.

3.2.6.3.1. Sélection ascendante

La sélection ascendante consiste à construire le meilleur modèle à une variable puis à ajouter une à une les variables afin d'améliorer les performances du modèle.

Supposons 3 variables X_1 , X_2 , X_3 telles que X_2 ait les meilleures performances seules et telles que l'ajout de la variable X_1 augmente plus les performances du modèle que l'ajout de la variable X_3 (Tableau 14).

Tableau 14 : Tableau illustrant la sélection de variables par la sélection ascendante.

Itération	Variables sélectionnées
Itération 1	X_2
Itération 2	X_2 et X_1
Itération 3	X_2 , X_1 et X_3

Les performances du modèle à chaque itération serviront comme critères de sélection des variables. Par exemple, si les performances du modèle à l'itération 2 et à l'itération 3 ne sont pas significativement différentes alors les variables X_2 et X_1 pourront être retenues.

3.2.6.3.2. Sélection descendante

La sélection descendante fonctionne dans le sens opposé à la méthode ascendante en partant d'un modèle avec toutes les variables puis le modèle élimine petit à petit la variable la moins significative, celle qui provoque une diminution de la métrique de performance.

Supposons les mêmes 3 variables que précédemment X_1 , X_2 , X_3 tel que X_2 ait les meilleures performances seules et tel que le retrait de la variable X_3 diminue moins les performances du modèle que le retrait de la variable X_1 (Tableau 15).

Tableau 15 : Tableau illustrant la sélection de variables par la sélection descendante.

Itération	Variables sélectionnées
Itération 1	X_2 , X_1 et X_3
Itération 2	X_2 et X_1
Itération 3	X_2

La sélection des variables se fait de la même manière que la sélection des variables en sélection ascendante bien que le départ soit différent.

3.2.6.3.3. Sélection *stepwise*

La sélection *stepwise* part du même principe que la sélection ascendante, la subtilité réside dans la possible élimination d'une variable introduite qui peut être éliminée lors des itérations suivantes car il peut arriver que l'introduction de certaines variables ne soient plus significatives après l'introduction d'autres variables.

En reprenant l'exemple de la partie sélection ascendante et une nouvelle variable X_4 telle que l'ajout de cette variable et la suppression de la variable X_3 donne un modèle plus performant qu'un modèle avec les deux variables (Tableau 16).

Tableau 16 : Tableau explicatif de la sélection de variables par la sélection *stepwise*.

Itération	Variables sélectionnées
Itération 1	X_2
Itération 2	X_2 et X_1
Itération 3	X_2 , X_1 et X_3
Itération 4	X_2 , X_1 et X_4

La sélection des variables se fait de la même manière que la sélection des variables en sélection ascendante et descendante bien que le départ soit différent.

3.3. Les modèles

3.3.1. La régression linéaire multiple

3.3.1.1. Principe

La régression linéaire multiple est une méthode mathématique permettant de modéliser le comportement d'une variable cible par une équation linéaire avec comme paramètres les variables étudiées pondérées par des coefficients.

Elle permet donc de construire une relation mathématique entre une variable cible et des variables explicatives⁴⁸.

Par exemple, il est possible de modéliser le prix d'une voiture à l'aide de différentes variables explicatives à savoir le fait qu'elle soit cylindrée, la puissance de la voiture, la longueur, la largeur, le poids et la vitesse de pointe de la voiture⁴⁸ :

$$\begin{aligned} & \text{Prix d'une voiture} \\ & = \\ & \beta_1 + \beta_2 * \text{cylindrée} + \beta_3 * \text{puissance} + \beta_4 * \text{longueur} + \beta_5 * \text{largeur} + \\ & \beta_6 * \text{poids} + \beta_7 * \text{vitesse de pointe} + \varepsilon \end{aligned}$$

3.3.1.2. Modélisation⁴⁹

Voici l'équation générale d'une régression linéaire multiple :

$$\hat{y} = \theta_0 + \sum_{i=1}^n \theta_i X_i + \varepsilon$$

Dans cette équation, \hat{y} est la variable réponse, c'est la variable cible. Les X_i sont les variables explicatives, celles qui permettent d'expliquer la variable réponse \hat{y} . Les coefficients θ_i devant les variables explicatives représentent les poids des variables qui sont le reflet de l'importance de la variable sur la variable cible. En effet, si θ est négatif cela signifie que la variable fait diminuer la variable cible. Si θ est positif, cela implique que la variable augmente la variable cible. Le θ_0 est un coefficient nommé intercept et indépendant des facteurs, c'est la valeur de la variable cible lorsque que toutes les variables explicatives sont nulles. Il y a également un terme d'erreur notée ε qui permet de mesurer l'écart entre les prédictions du modèle et les valeurs réelles. Ce terme doit être le plus faible possible.

En régression linéaire multiple, les variables explicatives peuvent également interagir entre elles. Dans ce cas, une nouvelle caractéristique est créée. Souvent cette caractéristique est égale au produit des deux variables qui interagissent entre elles.

Le but de la régression linéaire multiple est donc de déterminer les différents θ_i à l'aide de nos données. Voici la formule de détermination des coefficients θ_i par la méthode des moindres carrés qui est une méthode qui consiste à minimiser les résidus :

$$\theta = (X^T X)^{-1} X^T y$$

θ représente la matrice qui contient les coefficients de la régression linéaire multiple, la matrice X est la matrice qui contient les différentes valeurs des variables explicatives, la matrice X^T est la transposée de cette matrice la différence signifie que les lignes et les colonnes sont inversées. L'exposant -1 du produit matriciel de la matrice X par sa transposée représente l'inverse du résultat de ce produit matriciel. Enfin la matrice y est la matrice des observations de la variable à prédire.

3.3.2. La modélisation bayésienne

3.3.2.1. Principe³⁴

La modélisation bayésienne se base sur la notion de modèle populationnel décrit plus tard dans le manuscrit. Le but de la modélisation bayésienne est de construire un modèle de population avec les données existantes qui permettrait de prédire les caractéristiques pharmacocinétiques pour un nouvel individu grâce à ce modèle de population.

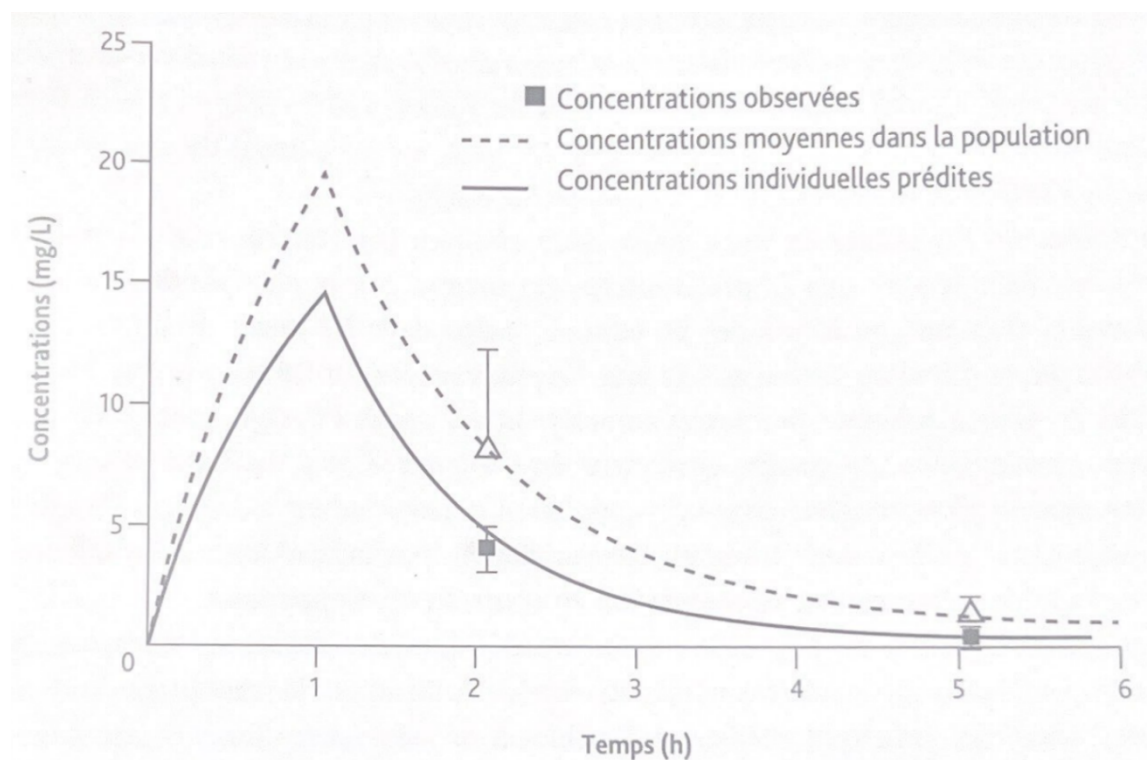


Figure 27 : Exemple d'utilisation de l'approche bayésienne pour la détermination des paramètres pharmacocinétiques d'un individu³⁴.

Ce graphique (Figure 27) montre en pointillés les concentrations moyennes de la population, en trait plein les concentrations individuelles prédites grâce au modèle de population et aux caractéristiques du patient, les carrés les concentrations moyennes observées dans la population aux temps t récoltés et les triangles les concentrations récoltées aux temps t . Le modèle de population permet de prédire le profil pharmacocinétique d'un individu même avec seulement deux concentrations.

3.3.2.2. Modélisation³¹

La modélisation populationnelle se fait en trois étapes³⁴ :

1. Définition du modèle de structure
2. Définition des modèles statistiques à savoir le modèle décrivant la variabilité interindividuelle et le modèle décrivant la variabilité résiduelle. Le modèle obtenu s'appelle le modèle pharmacostatistique.
3. Intégration au modèle de structure des équations mathématiques liant les covariables expliquant la variabilité aux paramètres pharmacocinétiques.

3.3.2.2.1. Modèle de structure

Le modèle de structure est un modèle qui décrit le mieux nos données en l'absence de covariables. Le modèle de structure doit être choisi entre les différents modèles de l'analyse compartimental dont les principaux sont :

- Le modèle mono-compartimental
- Le modèle bi-compartimental
- Le modèle tri-compartimental

Il existe en effet d'autres modèles avec plus de compartiments pour décrire la pharmacocinétique d'une substance, les principaux sont ceux énoncés au-dessus.

Un compartiment en pharmacocinétique représente un espace fictif de distribution où le médicament est considéré comme réparti de manière homogène. Dans cet espace, le principe actif arrive à la même vitesse, se distribue instantanément et ressort à la même vitesse en tout point du compartiment.

Un compartiment ne représente pas un tissu particulier mais un ensemble de tissus où la cinétique du principe actif est la même. Dans la modélisation compartimentale, plusieurs modèles peuvent co-exister (compartiment central et compartiments périphériques) mais il en faut toujours un, *a minima* le plasma, défini comme le compartiment central à partir duquel les variations quantitatives en principe actif sont mesurées.

Le modèle mathématique doit respecter la loi de conservation de masse ce qui implique de décrire les vitesses de transfert du principe actif entre les différents compartiments. La vitesse de transfert est une quantité de molécules par unités de temps. Ces vitesses sont définies par une constante notée k .

3.3.2.2.1.1. Modèle mono-compartmental

Le modèle mono-compartmental est défini par une fonction exponentielle unique qui décrit la décroissance des concentrations plasmatiques due à l'élimination du médicament (Figure 28). Lorsque le médicament est administré par voie bolus IV l'équation est :

$$C(t) = C_0 \cdot e^{-k_e t}$$

Avec C_0 la concentration plasmatique initiale et k_e la constante de vitesse d'élimination du médicament.

Lorsque le médicament est administré par voie extravasculaire, une première phase croissante est observée suivie d'une phase décroissante. L'équation décrivant ce cas est :

$$C(t) = -A \cdot e^{-k_a t} + B \cdot e^{-k_e t}$$

Avec A le paramètre de hauteur de la phase croissante, B le paramètre de hauteur de la phase décroissante et les constantes k_a et k_e les constantes de vitesse d'absorption et d'élimination respectivement.

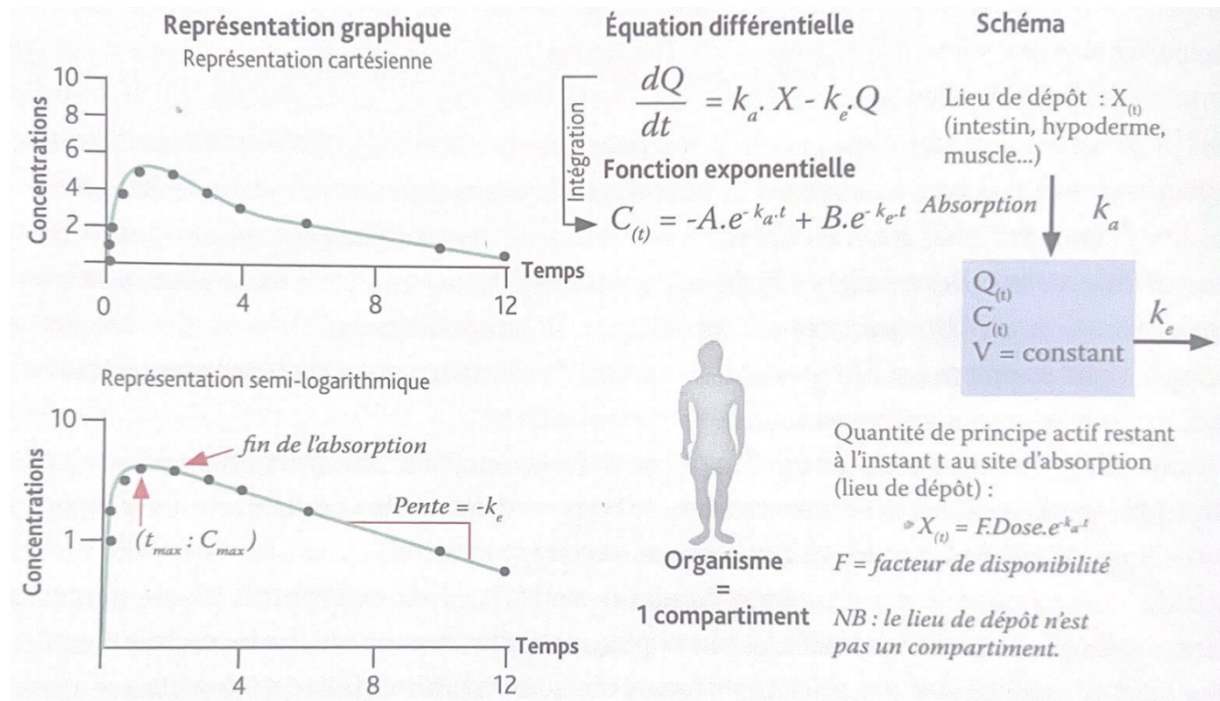


Figure 28 : Représentation graphique et mathématique du modèle à 1 compartiment avec administration extravasculaire³¹.

3.3.2.2.1.2. Modèle bi-compartmental

Le modèle bi-compartmental est représenté par deux compartiments : un compartiment central et un compartiment périphérique (Figure 29). Ce modèle est caractérisé par deux phases graphiques décroissantes.

Dans le cas d'une administration bolus IV l'équation s'écrit :

$$C(t) = A \cdot e^{-\alpha \cdot t} + B \cdot e^{-\beta \cdot t}$$

Avec α et β les pentes des phases croissantes et décroissantes respectivement.

Dans ce contexte, en notant le compartiment central comme le compartiment 1 et le compartiment périphérique comme le compartiment 2 il est possible de définir 3 constantes de vitesse : k_{12} , k_{10} et k_{21} .

La constante de vitesse de distribution du compartiment central vers le compartiment périphérique notée k_{12} qui se définit par l'équation suivante :

$$v_{12} = -k_{12} \cdot Q_1$$

Avec Q_1 la quantité de médicament dans le compartiment 1.

La constante de vitesse d'élimination notée k_{10} se définit par l'équation suivante :

$$v_{10} = -k_{10} \cdot Q_1$$

Une autre constante de vitesse doit être prise en considération qui va dans le sens inverse de la vitesse d'élimination, c'est la vitesse de retour du médicament du compartiment périphérique vers le compartiment central. Elle se définit de la façon suivante :

$$v_{21} = +k_{21} \cdot Q_2$$

Avec Q_2 la quantité de médicament dans le compartiment 2.

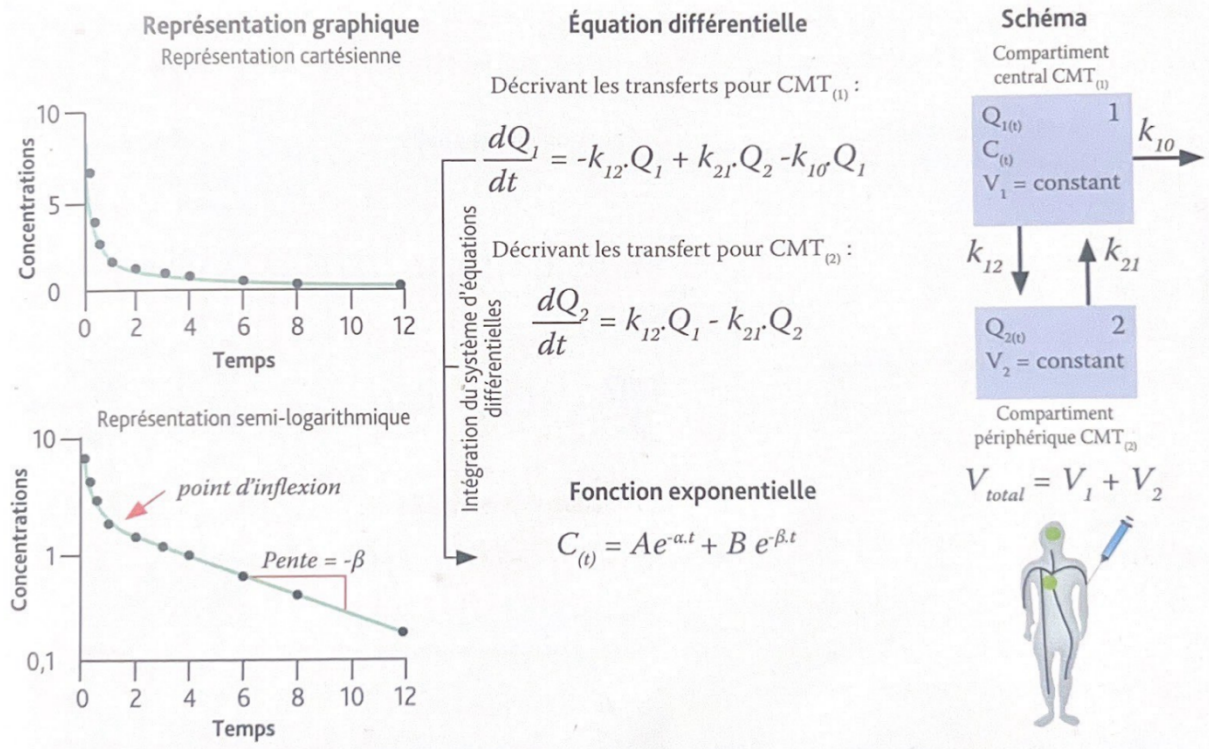


Figure 29 : Représentation graphique et mathématique du modèle à 2 compartiments avec administration bolus intravasculaire³¹.

Lorsque le médicament est administré par voie extravasculaire, l'équation décrivant l'évolution des concentrations en fonction du temps devient :

$$C(t) = -A \cdot e^{-\alpha \cdot t} + B \cdot e^{-\beta \cdot t} + C \cdot e^{-\gamma \cdot t}$$

Avec C et γ des paramètres de hauteur et de pente de la phase décroissante respectivement (Figure 30).

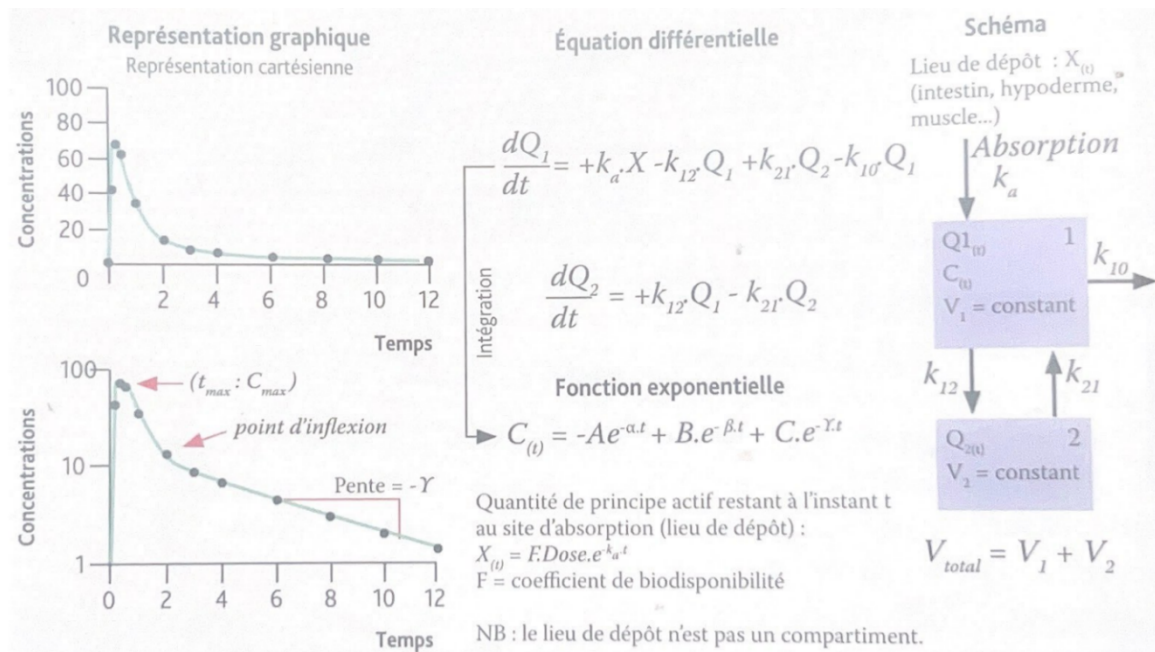


Figure 30 : Représentation graphique et mathématique du modèle à 2 compartiments avec administration extravasculaire³¹.

3.3.2.2.1.3. Modèle tri-compartimental

Dans le cadre du modèle tri-compartimental, l'organisme est représenté par 3 compartiments. Un compartiment central et 2 compartiments périphériques. Ce type de modèle est représenté graphiquement par trois phases décroissantes.

Le troisième compartiment correspond à un compartiment profond où la concentration du principe actif met plus de temps à s'équilibrer avec celle du compartiment central. La constante de vitesse de retour du compartiment périphérique vers le compartiment central est alors très faible.

Lorsque le médicament est administré en bolus, l'équation est :

$$C(t) = A.e^{-\alpha.t} + B.e^{-\beta.t} + C.e^{-\gamma.t}$$

3.3.2.2.2. Modèles statistiques³⁴

3.3.2.2.2.1. Modèle décrivant la variabilité interindividuelle

Dans ces modèles, il est considéré que les individus de la population n'ont pas tous les mêmes paramètres pharmacocinétiques. Partant de ce principe, les paramètres pharmacocinétiques individuels peuvent être considérés comme des variables aléatoires dont il faudra estimer la distribution en fonction des paramètres moyens de la population.

L'hypothèse la plus utilisée est que le logarithme du paramètre suit une loi normale. Le modèle est alors appelé modèle proportionnel dont l'équation est :

$$\text{Log}(P_i) = \text{Log}(\tilde{P}_i) + \eta_i$$

$$\tilde{P}_i = h(P, \text{COV}_i)$$

Avec P_i la valeur du paramètre pharmacocinétique chez l'individu, \tilde{P}_i la valeur moyenne de ce paramètre pharmacocinétique dans la population qui ont les mêmes valeurs de covariables, η_i l'écart à la médiane pour cet individu et h la fonction des effets fixes (paramètres dont la valeur ne varie pas d'un sujet à un autre) et des covariables COV_i .

Il existe deux grandes approches pour modéliser la distribution des η_i :

- L'approche paramétrique
- L'approche non paramétrique

Dans l'approche paramétrique, la distribution des η_i dans la population est supposée suivre une loi normale centrée sur 0 et de variance ω^2 : $\eta_i \sim N(0, \omega^2)$. Tandis que dans l'approche non paramétrique, aucune hypothèse ne sera formulée sur la distribution des η_i dans la population.

3.3.2.2.2. Modèle décrivant la variabilité résiduelle

Les concentrations prédites par le modèle et les concentrations observées peuvent ne pas être égales comme c'est le cas la plupart du temps. Cet écart entre les concentrations prédites et les concentrations observées est appelé la variabilité résiduelle, elle se note ε .

Plusieurs modèles d'erreurs existent pour décrire la variabilité résiduelle :

- Le modèle additif qui part du principe que l'incertitude est indépendante du niveau de concentration il se définit comme :

$$DV = IPRED + \varepsilon$$

- Le modèle proportionnel qui part du principe que l'incertitude est proportionnelle au niveau des concentrations il se définit comme :

$$DV = IPRED + \varepsilon \cdot IPRED$$

- Et le modèle mixte qui est un mélange du modèle proportionnel et du modèle additif il se définit comme :

$$DV = IPRED + \varepsilon_1.IPRED + \varepsilon_2$$

Avec IPRED (*Individual predicted concentrations*), les prédictions individuelles des concentrations.

La variabilité résiduelle est la somme des erreurs associés au résultat final, comme les erreurs analytiques lors de la détermination des concentrations plasmatiques ou encore les erreurs liées aux relevés des horaires de prélèvements.

Une autre source de variabilité résiduelle peut être liée à un défaut de choix du modèle de structure peu adapté aux données. En effet, si la variabilité résiduelle est supérieure à l'incertitude analytique alors il faut envisager de changer de modèle structural pour avoir une variabilité résiduelle plus faible et plus acceptable.

3.3.2.2.3. Recherche de covariables

La recherche de covariables est la troisième étape de la modélisation d'une populationnelle. Elle permet d'identifier les relations entre les covariables et les paramètres pharmacocinétiques. Ces relations peuvent être linéaires ou allométriques.

Dans le cas d'une relation linéaire, la relation entre les paramètres pharmacocinétiques et les covariables peut s'écrire de la façon suivante :

$$\tilde{P}_i = a.COV_i + b$$

Avec a et b les paramètres à optimiser.

Dans le cas d'une relation allométrique (relation non-linéaire), la relation entre les paramètres pharmacocinétiques et les covariables pourrait s'écrire de la façon suivante :

$$\tilde{P}_i = a.COV_i^b$$

3.3.2.2.4. Estimation des paramètres

Un modèle de population peut être défini comme un modèle non linéaire à effets mixtes. Il est qualifié de non-linéaire car la relation entre les concentrations et les paramètres est une relation non linéaire. Le modèle de population est aussi un modèle à effets mixtes car il contient deux types d'effets à savoir les effets fixes et les effets aléatoires.

Les effets fixes sont les paramètres dont la valeur est invariable d'un individu à un autre tandis que les effets aléatoires sont des paramètres qui varie d'un sujet à un autre.

Dans le cas d'un modèle de population, les paramètres à estimer sont donc les effets fixes et la variance.

Dans le cas d'un modèle sans covariables, les effets fixes représentent la moyenne de chacun des paramètres pharmacocinétiques (PK). Pour estimer ces effets, il faut se baser sur le principe du maximum de vraisemblance, une technique permettant de trouver les effets les plus probables correspondant aux observations.

Cependant il n'existe pas de méthode pour calculer de manière exacte la fonction vraisemblance d'un modèle non-linéaire, pour la déterminer des itérations successives sont effectuées : à chaque pas, une combinaison différente des valeurs de tous les paramètres pharmacocinétiques est estimée. La méthode d'itération a un impact sur la qualité des données.

Chaque combinaison de paramètres correspondra donc à une probabilité d'observation des données observées. L'ensemble de ces probabilités forment la fonction de vraisemblance.

En pratique, une approximation de la fonction de vraisemblance est utilisée en calculant la fonction objective (OFV) qui se calcule de la façon suivante :

$$OFV = -2 \times \ln (Vraisemblance)$$

Pour maximiser la vraisemblance, il faut donc que la différence entre les concentrations moyennes prédites et les concentrations observées soit la plus faible possible.

Dans le cadre de la LSS, les concentrations observées les plus informatives pour le calcul du profil pharmacocinétique de la substance étudiée sont sélectionnées. Ces concentrations vont servir à déterminer le modèle de population permettant de calculer les paramètres pharmacocinétiques d'un individu à l'aide d'une formule reliant les concentrations aux paramètres pharmacocinétiques.

Pour la prédiction, de nouvelles concentrations seront à considérer et permettront de calculer les paramètres pharmacocinétiques d'un individu spécifique. Ces paramètres seront ensuite utilisés pour retracer le profil pharmacocinétique complet de la substance étudiée pour un individu spécifique.

Ces concentrations seront ensuite intégrées avec les autres données afin de modifier la formule d'estimation des paramètres pharmacocinétiques avec de nouvelles concentrations³⁷.

3.3.3. Comparaison entre la régression linéaire multiple et la modélisation bayésienne³⁷

L'approche bayésienne est différente de la régression linéaire (Tableau 17). La régression linéaire se base sur l'exploitation des observations tandis que l'approche bayésienne se base sur des connaissances qui seront affinées grâce aux observations.

Concernant les avantages de l'approche bayésienne, l'approche bayésienne a plus de flexibilité dans les prélèvements des concentrations car il est possible de les prélever à des moments différents. L'inconvénient majeur de l'approche bayésienne est qu'elle nécessite l'utilisation de logiciels complexes.

La régression linéaire multiple a l'avantage de pouvoir estimer l'AUC uniquement à l'aide d'une formule reliant l'AUC à différentes variables explicatives dont les concentrations mesurées, et de pouvoir se faire avec des logiciels relativement simples. Cependant elle présente aussi des inconvénients. Elle est moins flexible en termes de temps car les échantillons doivent être prélevés selon un calendrier défini, à des moments déjà définis.

Tableau 17 : Tableau comparatif de la régression linéaire multiple et de la modélisation bayésienne.

	Régression linéaire multiple	Modélisation bayésienne
Principe	Exploitation des données	Exploitation des connaissances et des observations
Flexibilité	-	+
Facilité	+	-

3.4. La validation des modèles

3.4.1. La validation de la régression linéaire multiple

3.4.1.1. Validation par les métriques

3.4.1.1.1. Le coefficient de détermination (r^2)

Le coefficient de détermination noté de manière conventionnel r^2 est un coefficient qui mesure la force d'adéquation entre le modèle et les données disponibles. Un modèle qui a un r^2 de 0.95 signifie que ce modèle permet d'expliquer 95% de la variabilité des données.

Le r^2 est un coefficient qui varie de 0 à 1. Plus le coefficient est proche de 1 plus le modèle explique correctement nos données. La formule du coefficient de détermination⁴⁸ est la suivante :

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Avec \hat{y}_i les prédictions, y_i les valeurs réellement observées et \bar{y} les valeurs moyennes de y_i .

3.4.1.1.2. Le coefficient de détermination ajusté ($r^2_{\text{Ajusté}}$)

Le $r^2_{\text{Ajusté}}$ s'interprète de la même façon que le r^2 . La différence réside dans le calcul. Le $r^2_{\text{Ajusté}}$ prend en compte le nombre de variables et le nombre d'observations⁴⁸ ce qui n'est pas le cas pour le r^2 .

$$r^2_{\text{ajusté}} = \frac{(n - 1) * r^2 - p}{n - p - 1}$$

Avec n et p respectivement le nombre d'observations et le nombre de variables.

3.4.1.1.3. La Root Mean Squared Error⁵¹

La RMSE nommée aussi erreur quadratique moyenne représente les erreurs de prédictions. C'est le reflet de la précision du modèle. Elle permet d'évaluer à quel point les données s'éloignent de la courbe de régression. Plus elle est faible moins il y a d'erreur et plus la courbe de régression décrit correctement nos données et inversement. La formule de la RMSE est la suivante :

$$RMSE(\%) = 100 \cdot \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N \left(\frac{\hat{y}_i - y_i}{y_i} \right)^2}$$

3.4.1.1.4. La Mean Predictive Error⁵¹

La MPE est différente de la RMSE mais complémentaire de cette métrique. Elle compare les prédictions aux valeurs réellement observées. Elle représente le biais du modèle. C'est l'erreur moyenne en pourcentage. Ce paramètre à l'avantage de permettre la comparaison de données de différentes échelles.

$$MPE(\%) = \frac{100}{N} \cdot \sum_{i=1}^N \left(\frac{\hat{y}_i - y_i}{y_i} \right)$$

3.4.1.1.5. La métrique J

La métrique J (comme Jean-Luc) est une métrique qui n'a pas encore été décrite dans la littérature et qui permet de comparer les modèles entre eux. La construction de cette métrique se base sur les intervalles de définition du r^2 , $r^2_{\text{ajusté}}$, de la $RMSE_{\% \text{Décimale}}$ et de la $|MPE|_{\% \text{Décimale}}$:

$$\begin{aligned} 0 &\leq r^2 \leq 1 \\ 0 &\leq r^2_{\text{ajusté}} \leq 1 \\ 0 &\leq RMSE_{\% \text{Décimale}} \leq 1 \\ 0 &\leq |MPE|_{\% \text{Décimale}} \leq 1 \end{aligned}$$

Il y a une nécessité de modifier les intervalles pour avoir un sens commun à toutes ces métriques car l'augmentation du r^2 et du $r^2_{\text{ajusté}}$ indique de bonnes performances tandis que la diminution de la $RMSE_{\% \text{ décimale}}$ et de la $|MPE|_{\% \text{Décimale}}$ indique de bonnes performances. Il faut donc choisir un intervalle commun à ces différentes métriques. Pour cela, il faut modifier les intervalles de façon à ce que l'augmentation ou la diminution de toutes les métriques indique de meilleures performances du modèle. La diminution des métriques redéfinies a été choisie de manière arbitraire. Les intervalles ont donc été modifiés de la façon suivante :

$$\begin{aligned} 0 &\leq 1 - r^2 \leq 1 \\ 0 &\leq 1 - r^2_{\text{ajusté}} \leq 1 \\ 0 &\leq RMSE_{\% \text{Décimale}} \leq 1 \\ 0 &\leq |MPE|_{\% \text{Décimale}} \leq 1 \end{aligned}$$

Le paramètre J est alors défini comme la somme des nouvelles métriques : **$J = 2 - r^2 - r^2_{\text{ajusté}} + RMSE_{\% \text{ décimale}} + |MPE|_{\% \text{décimale}}$** . Ce paramètre est donc compris entre 0 et 4 par addition des bornes des intervalles des métriques redéfinies.

Pour pouvoir appliquer cette métrique, il faut que le r^2 et le $r^2_{\text{ajusté}}$ soit supérieur à 0. Plus un modèle a une valeur J proche de 0, meilleur sera le modèle. De plus, il ne faut pas oublier que cette métrique a pour vocation de comparer les modèles entre eux mais pas de les valider car un modèle peut très bien avoir un J très proche de 0 sans remplir les conditions sur les métriques lorsqu'il y en a. Il faut donc un modèle ayant une métrique J la plus proche de 0 tout en respectant les conditions définies sur les métriques utilisées de manière individuelles.

Par exemple, voici deux modèles : un modèle A ayant un $r^2 = 0.89$, un $r^2_{\text{ajusté}} = 0.79$, une $\text{RMSE}_{\% \text{Décimale}} = 0$ et une $|\text{MPE}|_{\% \text{Décimale}} = 0$ et un modèle B ayant un $r^2 = 0.9$, un $r^2_{\text{ajusté}} = 0.8$, une $\text{RMSE}_{\% \text{Décimale}} = 0.20$ et une $|\text{MPE}|_{\% \text{Décimale}} = 0.20$.

Dans ce cas, le meilleur modèle est défini par un $r^2 > 0.9$, un $r^2_{\text{ajusté}} > 0.8$, une $\text{RMSE}_{\% \text{Décimale}} < 20$ et une $|\text{MPE}|_{\% \text{Décimale}} < 20$.

En calculant les métriques J pour ces deux modèles les valeurs obtenues sont $J_A = 0.32$ et $J_B = 0.7$. Le modèle A a une métrique J plus proche de 0 que le modèle B ce qui indique qu'en moyenne, le modèle A serait meilleur que le modèle B. Cependant, le modèle A ne satisfait pas les valeurs requises des différentes métriques définies alors que le modèle B les respecte. Dans cet exemple, le modèle B sera donc retenu.

3.4.1.2. Validation croisée⁵²

La validation croisée est une technique qui permet de lutter contre le surapprentissage du modèle. Le surapprentissage désigne le fait que le modèle est très performant sur les données qui lui ont servi d'entraînement au point de perdre sa capacité à généraliser correctement sur de nouvelles données.

Le surapprentissage a plusieurs origines dont les plus fréquentes sont une complexité trop importante du modèle, un nombre de données trop faible et pas assez représentatif ou encore un nombre de variables trop important.

Pour lutter contre le surapprentissage, un des moyens qui peut être mis en place est de réaliser une validation croisée.

La validation croisée est une procédure de rééchantillonnage et consiste à séparer les données en deux ou plusieurs sous parties. Dans le cas de deux parties, il y a une partie appelée données d'entraînement qui sert à trouver les paramètres du modèle et une partie appelée données de test utilisée pour simuler une situation réelle sur des données non exploitées pour l'entraînement du modèle.

Plusieurs techniques de validation croisée existent, les plus utilisées sont la validation croisée *K-folds* et la validation croisée *leave-one-out* (LOOCV).

3.4.1.2.1. Cross validation *K-folds*

La validation croisée *K-folds* est une technique qui permet de séparer les données en K sous-ensembles appelés *folds*. À chaque itération, un *fold* différent est utilisé pour l'entraînement du modèle et les autres *folds* en données de test pour l'évaluation des performances du modèle. Le processus s'arrête lorsque tous les *folds* ont été utilisés une fois comme entraînement.

La performance du modèle est donc la moyenne des performances du modèle à chaque itération. Cette technique est utilisée la plupart du temps lorsque les données sont nombreuses car elle est moins coûteuse en temps de calcul.

Par exemple, voici un jeu de données comportant 1000 données découpé en 4 *fold*s qui contiennent donc chacun 250 données (Tableau 18). Un exemple de performance en validation croisée *K-folds* d'un modèle de régression linéaire multiple (la MPE dans cet exemple) serait le suivant :

Tableau 18 : Tableau explicatif de la validation croisée K-folds.

Modèles	Fold 1	Fold 2	Fold 3	Fold 4	MPE (%)
Modèle 1	E	T	T	T	0,83
Modèle 2	T	E	T	T	0,9
Modèle 3	T	T	E	T	0,85
Modèle 4	T	T	T	E	0,89

Avec E pour entraînement et T pour test.

Dans cet exemple, la performance du modèle de régression linéaire multiple est donc égale à :

$$MPE_{moyenne} = \frac{0,83 + 0,9 + 0,85 + 0,89}{4} = 0.8675 \%$$

La *cross validation K fold* permet également de calculer des intervalles de confiance des coefficients du modèle (c'est-à-dire des limites supérieures et inférieures pour chaque coefficient) en récoltant les coefficients de chaque variable à chaque itération.

3.4.1.2.2. **Leave one out cross validation (LOOCV)**

La validation *LOOCV* est une technique de validation croisée *K folds* avec un nombre de *fold*s égal à n. Dans ce cas particulier, chaque observation sera utilisée une seule et unique fois comme donnée de test à chaque itération.

La performance moyenne du modèle sera donc la moyenne des performances du modèle à chaque itération comme en validation croisée *K folds*. Cette technique est plutôt utilisée lorsque le nombre de données est trop faible pour faire une *cross validation K-folds*. Elle a l'avantage d'être très puissante car elle évalue le modèle sur toutes les données disponibles.

Par exemple, voici un jeu de données comportant 10 données. Un exemple d'évaluation de performance en validation croisée *LOOCV* d'un modèle de régression linéaire multiple (la MPE dans cet exemple) serait le suivant (Tableau 19) :

Tableau 19 : Tableau explicatif de la validation croisée *LOOCV*.

Modèles	D ₁	D ₂	D ₃	D ₄	D ₅	D ₆	D ₇	D ₈	D ₉	D ₁₀	MPE (%)
M ₁	T	E	E	E	E	E	E	E	E	E	0,7
M ₂	E	T	E	E	E	E	E	E	E	E	0,8
M ₃	E	E	T	E	E	E	E	E	E	E	0,73
M ₄	E	E	E	T	E	E	E	E	E	E	0,77
M ₅	E	E	E	E	T	E	E	E	E	E	0,82
M ₆	E	E	E	E	E	T	E	E	E	E	0,71
M ₇	E	E	E	E	E	E	T	E	E	E	0,73
M ₈	E	E	E	E	E	E	E	T	E	E	0,75
M ₉	E	E	E	E	E	E	E	E	T	E	0,79
M ₁₀	E	E	E	E	E	E	E	E	E	T	0,81

Avec M_n les différents modèles, et D_n les différentes données.

Dans cet exemple, la performance du modèle de régression linéaire multiple est donc égale à :

$$MPE_{moyenne} = \frac{0,7 + 0,8 + 0,73 + 0,77 + 0,82 + 0,71 + 0,73 + 0,75 + 0,79 + 0,81}{10} = 0,761$$

La cross validation *LOOCV* permet également de calculer des intervalles de confiance des coefficients du modèle.

3.4.1.3. Validation des hypothèses

3.4.1.3.1. Graphique des valeurs prédites en fonction des valeurs observées (*predicted vs observed*)

Le *predicted vs observed* est un graphique qui met en évidence l'écart entre les prédictions et les observations.

En abscisse de ce graphique (Figure 31) se trouvent les valeurs réelles et en ordonné les prédictions. Pour comparer plus facilement, la droite d'identité est tracée. Les points qui sont sur la droite représentent les valeurs qui ont été correctement prédites. Plus les points sont éloignés de cette droite plus le modèle s'est éloigné dans sa prédiction.

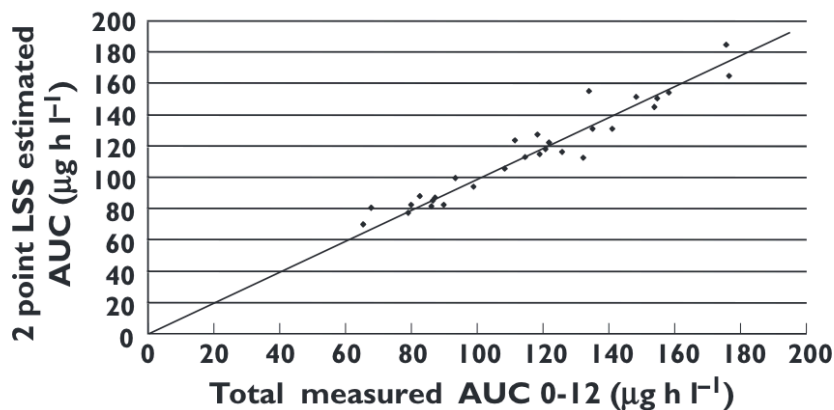


Figure 31 : Exemple de *predicted vs observed*⁵³.

3.4.1.3.2. Linéarité du modèle

L'hypothèse de linéarité traduit le fait qu'il existe une relation linéaire entre la variable à prédire et ses variables explicatives. Elle se vérifie graphiquement en regardant le graphique qui représente les résidus « *studentisés* » en fonction des prédictions.

Les résidus « *studentisés* » sont les résidus du modèle ajustés pour suivre une loi de Student. Pour que cette hypothèse soit vérifiée, il faut que ces résidus soit compris entre -3 et 3 qui représente les limites à 99% d'une loi normale⁵⁴.

3.4.1.3.3. Homoscédasticité des résidus⁵⁵

L'homoscédasticité représente le fait que les résidus du modèle ont une variance constante et qu'il n'y a pas de tendance particulière dans la distribution de ces derniers.

L'évaluation repose généralement sur l'analyse graphique, si une tendance apparaît l'homoscédasticité risque de ne pas être vérifiée (exemple avec la Figure 32).

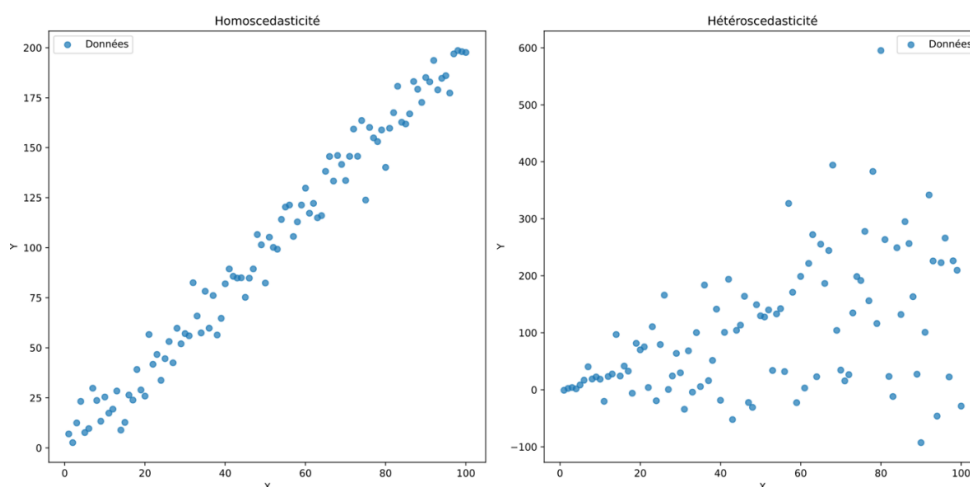


Figure 32 : Exemple de graphique illustrant l'homoscédasticité (à droite) et l'hétéroscedasticité (à gauche).

L'homoscédasticité peut également se vérifier à l'aide du test statistique de Breusch-Pagan en faisant l'hypothèse initiale nulle H_0 que les données sont homoscédastiques. Si la p -value de la statistique de test est supérieure au niveau de risque α choisi alors l'hypothèse nulle H_0 ne peut pas être rejetée. Ce qui signifie qu'il n'y a pas d'hétéroscédasticité détectée dans les résidus.

3.4.1.3.4. Normalité des résidus

L'hypothèse de normalité des résidus repose sur le fait que leur distribution suit une loi normale. Cette hypothèse est importante car le fait que les résidus suivent une loi normale permet d'obtenir des prédictions fiables et sans biais. La normalité des résidus peut se vérifier à l'aide de tests statistiques comme le test de normalité de Shapiro-Wilk dont l'hypothèse initiale nulle H_0 estime que les données suivent une loi normale au risque α . Si la probabilité de la statistique de test est inférieure au risque α alors l'hypothèse nulle H_0 est rejetée⁵⁶.

3.4.1.3.5. Non-autocorrélation des résidus

L'hypothèse de non-autocorrélation des résidus permet de vérifier qu'il n'y a pas de corrélation entre les résidus c'est-à-dire qu'ils sont indépendants entre eux. La non-autocorrélation des résidus peut se vérifier à l'aide du test statistique de Durbin-Watson dont l'hypothèse nulle H_0 est basée sur l'absence d'autocorrélation des résidus. Une valeur de la statistique de test proche de 2 permet d'accepter l'hypothèse H_0 ⁵⁷.

3.4.1.3.6. Non multicolinéarité entre les variables

La multicolinéarité est un problème majeur en régression linéaire multiple. En effet, des variables fortement corrélées entre elles entraînent des variances très élevées dans les coefficients des modèles de régression linéaire multiple, ce qui fait que les estimations de ces coefficients deviennent imprécises.

Pour vérifier qu'il n'y a pas de multicolinéarité, il faut calculer le facteur d'inflation de la variance (VIF) qui est un indicateur global de multicolinéarité. Il se calcule de la façon suivante :

$$VIF_j = \frac{1}{1 - R_j^2}$$

Avec R_j^2 le coefficient de corrélation multiple entre la variable x_j et les autres variables explicatives au carré. Le coefficient de corrélation multiple est la valeur maximale prise par les coefficients de corrélation linéaire entre la variable x_1 et une combinaison linéaire des autres variables explicatives⁴⁸. Le R_j^2 est compris entre 0 et 1 ce qui implique que le VIF est compris entre 1 et $+\infty$. Plus le R_j^2 est proche de 0 plus le VIF approche la valeur 1 et moins les variables sont colinéaires entre elles.

3.4.1.3.7. Bland-Altman

Le graphique Bland-Altman permet de comparer deux techniques d'analyse entre elles. Il permet de comparer une technique de référence à une autre technique afin d'évaluer si les deux donnent des résultats similaires.

Ce graphique représente la différence entre les valeurs mesurées par les deux techniques en fonction de la moyenne des valeurs mesurées avec les deux techniques.

Ce graphique (Figure 33) est délimité par une limite supérieure et une limite inférieure qui représente les intervalles de confiance des valeurs. Il est également possible de tracer une ligne horizontale représentant le biais des techniques étudiées. Plus cette ligne est proche de 0 plus le biais est faible et donc plus les deux techniques donnent des valeurs qui sont relativement proches.

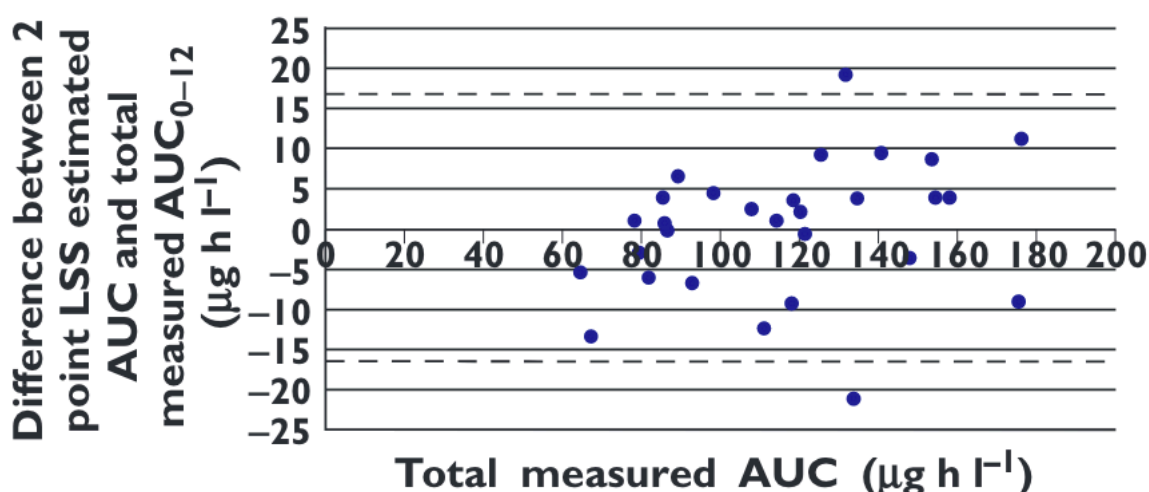


Figure 33 : Exemple de Bland-Altman⁵⁸.

3.4.1.3.8. Significativité des variables et du modèle

Il est également important de calculer la significativité du modèle et des variables explicatives du modèle. Cette significativité est calculée à l'aide de test d'hypothèse et se fonde sur l'interprétation de la *p-value*. Une *p-value* inférieure à 5% indique que le modèle et/ou les variables sont considéré(es) comme significatif(-ives)⁵⁹.

3.4.2. La validation de la modélisation bayésienne

3.4.2.1. Validation par les métriques

Les métriques utilisées pour évaluer une régression linéaire multiple définies ci-dessus peuvent également être utilisées pour la modélisation bayésienne.

Néanmoins, une autre métrique est très utilisée en modélisation bayésienne: la *Relative Standard Error* (RSE) qui permet d'évaluer la fiabilité du modèle et se définit par la formule suivante⁶⁰:

$$RSE(\%) = 100 \times \frac{\textit{Standard Error}}{\textit{estimated parameter value}}$$

Avec *Standard Error* qui signifie erreur standard représentant l'incertitude des paramètres estimés dans la population⁶⁰.

3.4.2.2. Validation par le *bootstrap*⁶¹

Une autre méthode de validation utilisée pour l'approche bayésienne est le *bootstrap*.

Le *bootstrap* est une technique de rééchantillonnage. Le fonctionnement est le suivant : premièrement, des tirages au sort au hasard avec remise sont effectués dans l'échantillon principal afin de construire des sous-échantillons (Tableau 20). Chaque sous-échantillon reconstruit permet de recalculer sur chacun d'entre eux les paramètres du modèle et de calculer à la fin un intervalle de confiance des paramètres de ce modèle à l'aide des paramètres obtenus sur chaque sous-échantillon. Cette méthode permet d'étudier la variabilité des paramètres du modèle.

Le nombre de simulation *bootstrap* doit être défini. Plus ce nombre est important, plus les estimations seront fiables mais plus il faudra de temps de calcul pour obtenir les résultats.

Tableau 20 : Tableau illustrant le *bootstrap*.

Échantillon initial	Sous - échantillon 1 bootstrap	Sous – échantillon 2 bootstrap	Sous – échantillon 3 bootstrap	Sous – échantillon 4 bootstrap
D ₁	D ₁	D ₂	D ₅	D ₁
D ₂	D ₁	D ₄	D ₂	D ₂
D ₃	D ₃	D ₄	D ₂	D ₁
D ₄	D ₅	D ₁	D ₅	D ₁
D ₅	D ₃	D ₅	D ₁	D ₃

À travers cet exemple, chaque échantillon servira pour calculer les paramètres du modèle. Les différents paramètres trouvés permettront de calculer un intervalle de confiance de ces paramètres.

3.4.2.3. Validation graphique

Dans le cadre de la validation graphique en approche bayésienne, plusieurs représentations existent : les graphiques *predicted vs observed* ou encore

les graphiques de la distribution des résidus (graphiques décrits dans la partie validation des hypothèses).

L'une des représentations graphiques la plus utilisée en modélisation bayésienne est le *Visual Predictive Check* (VPC). Elle permet d'évaluer la capacité prédictive d'un modèle. C'est une technique qui se base sur de multiples simulations à partir de la PKPOP et des données d'origine³⁴. La Figure 34 représente un exemple de VPC. La ligne rouge représente la concentration médiane observée, la zone rouge de délimitation représente l'intervalle de confiance à 95% obtenu par simulation de cette médiane. Cette zone est délimitée par des lignes rouges en pointillés qui représentent les percentiles observés à 5% et à 95%. Enfin les zones bleues représentent les intervalles de confiance à 95% pour les percentiles prédits qui sont les valeurs estimées divisant une distribution prédite en cent groupes égaux⁶².

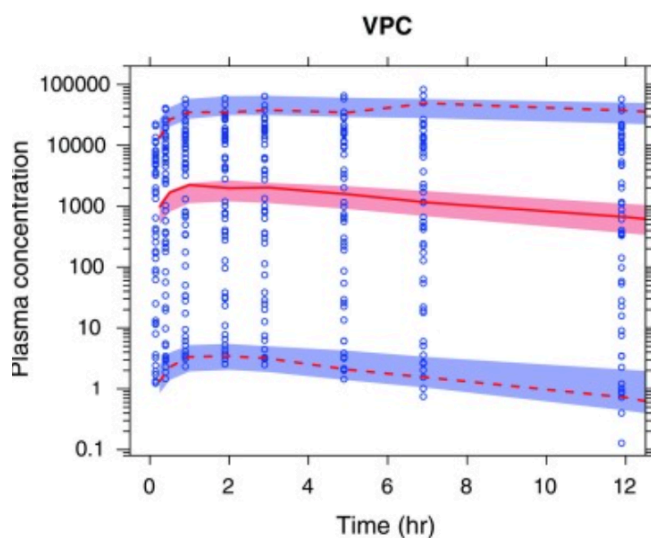


Figure 34 : Exemple de *Visual Predictive Check* (VPC)⁶².

3.5. Quelques exemples d'application médicales

La LSS est une technique très utilisée dans le domaine des essais cliniques et qui a fait l'objet de plusieurs publications, dans le domaine de la santé. Elle a été notamment utilisée pour :

- Prédire l'aire sous la courbe de la cyclosporine en trois points après l'administration de Neoral® (voie orale) et de Sandimmun® (voie IV)⁶³.
- Pour le suivi thérapeutique de la polymyxine B chez les patients chinois souffrant d'infections bactériennes à Gram négatif multirésistantes (voie IV)⁶⁴.
- Pour la prédiction de l'exposition systémique au nouvel agent anticancéreux E7070 (N-(3-chloro-7-indolyl)-1,4-benzenedisulphonamide, voie IV)⁶⁵.

- Pour l'individualisation de la dose de busulfan IV chez les enfants subissant une greffe de cellules souches hématopoïétiques (voie IV)⁶⁶.
- Pour la prédiction précise de la pharmacocinétique du Saroglitazar grâce à 3 points (voie orale)⁶⁷.
- Pour la détermination précise de la pharmacocinétique du facteur VIII à demi-vie prolongée chez les patients atteints d'hémophilie A sévère (voie IV)⁶⁸.
- Pour l'évaluation de la pharmacocinétique de l'acide mycophénolique chez les enfants atteints du syndrome néphrotique (voie orale)⁶⁹.
- Pour la surveillance thérapeutique du mycophénolate mofétil chez des patients japonais atteints de néphrite lupique (voie orale)⁷⁰.

Les applications médicales listées ci-dessus sont non exhaustives et démontrent que les applications de la LSS sont diverses et variées.

4. La LSS appliquée aux LNCs

4.1. Objectifs

Le but de ce travail est de réaliser une LSS avec 3 temps de prélèvements de sang maximum chez le rat pour des NCL en utilisant la régression linéaire multiple. En effet, la LSS permet de trouver les temps de prélèvement les plus informatifs et les covariables les plus informatives pour expliquer l'exposition (AUC) des rats aux NCL. Elle permet donc de réduire le nombre de prélèvements à réaliser pour obtenir le profil pharmacocinétique de ces dernières.

La réduction du nombre de prélèvements est souhaitée pour plusieurs raisons. Premièrement, elle permet de réduire les variations du volume de distribution qui ont lieu à chaque prélèvement. Car en effet, des prélèvements nombreux et volumineux peuvent diminuer le volume de distribution entraînant une modification de la clairance et donc un biais dans l'élimination de la molécule d'intérêt et de l'analyse pharmacocinétique. Elle permet aussi de gagner du temps et de diminuer les coûts d'analyse car les prélèvements, ainsi que leur analyse, peuvent s'avérer chronophages et coûteux. Enfin, la LSS permet aussi de réduire la potentielle souffrance animale liée à la prise de sang que les rats peuvent subir à chaque prélèvement. La réduction du nombre de prélèvements permet donc de diminuer le stress animal qui peut provoquer un biais dans les résultats d'analyse. De plus, augmenter le bien-être animal est une obligation éthique et légale.

La pharmacocinétique des NCL a été étudiée chez des rats de type Sprague Dawley après administration intraveineuse des NCL décrites dans l'étude de Lebreton et al.³⁵. Puis ces NCL ont été dosées à l'aide de la technique de dosage FRET qui permet de suivre leur évolution dans l'organisme en

fonction du temps. Jusqu'ici, aucune étude n'avait utilisé la méthode FRET pour étudier la pharmacocinétique des NCL^{71,35}.

4.2. Matériels et Méthodes

Les données analysées dans cette l'étude proviennent du travail réalisé par Lebreton *et al*³⁵ détaillé dans la partie « Intérêt de la technique FRET pour l'évaluation pharmacocinétique des NCL ». Ces données incluent les concentrations plasmatiques des différentes NCL à partir des prélèvements de sang, la dose injectée, la composition des NCL, le potentiel zêta des NCL, le poids du rat ainsi que leur âge. Les compositions de Lipoïd[®], de NaCl et d'eau purifiée n'ont pas été incluses car elles sont identiques peu importe le type de NCL ce qui n'apporterait pas d'informations supplémentaires pour le modèle. Au final, les variables explicatives conservées sont : les différentes concentrations plasmatiques aux différents temps de prélèvements : 5, 30, 60, 90, 120, 180, 240, 360, 600 et 1440 minutes ainsi que la concentration initiale C0 (à 0 minutes); le poids et l'âge des rats ; la dose injectée ; la quantité de Captex[®] Dil et de captex DiD ; la quantité de Kolliphor[®] ; la quantité d'eau purifiée ; la quantité de PEG, la quantité de SA ; la taille des NCL ainsi que leur potentiel zêta (jeu de données référencé en Annexe 4).

Ces données ont permis de calculer les AUC selon une approche non compartimentale de type log trapézoïdale à l'aide de PKanalix (lixfort version 2021, France). Cette étude a également été réalisée avec le langage de programmation Python *via* l'environnement de développement *Spyder* (version 5.4.2) pour réaliser la LSS. Les codes de l'étude sont donnés dans les Annexes 5 à 17.

À partir du jeu de données initial, de nouveaux jeux de données ont été créés en fonction du type de LSS souhaité. Ce qui revient à dire que des jeux de données spécifiques à la LSS 1 point, à la LSS 2 points et à la LSS 3 points ont été développés.

Pour la LSS à 1 point, un seul temps de prélèvement sera utilisé dans chaque jeu de données ; deux temps de prélèvements pour la LSS à 2 points et trois temps de prélèvements pour la LSS à 3 points.

De plus, de nouvelles variables ont été créés et incorporées dans chaque jeu de données lorsque c'est possible. Les variables créés sont les suivantes :

- Les interactions des concentrations (ex : $C5 * C360$)
- Les logarithmes des concentrations (ex : $\ln(C5)$; $\ln(C360)$)
- Les interactions entre les logarithmes des concentrations
(ex : $\ln(C5) * \ln(C360)$)

Cependant, les interactions entre les concentrations plasmatiques et les autres covariables n'ont pas été prises en compte pour que l'explicabilité du modèle ne soit pas trop complexe. En effet, il serait difficile d'interpréter la variable Captex Dil x C600. Les interactions entre les concentrations ont donc été les seules prises en compte.

Le logarithme des concentrations a également été construit et se justifie par le fait que le calcul de l'AUC a été réalisé selon l'approche log trapézoïdal qui prend en compte le logarithme des concentrations dans son équation.

Par type de LSS, toutes les combinaisons de jeux de données à 1, 2 ou 3 concentrations ont été créés. Les sous jeux de données contiennent alors :

- Les combinaisons des concentrations
- Les transformations de ces concentrations
- Les caractéristiques physico-chimiques des NCL
- Les caractéristiques individuelles des rats

Le nombre de jeux de données par type de LSS peut être calculé à l'aide du coefficient binomial, notons le N. Pour la LSS à 1 point $n=11$ et $k=1$, pour la LSS à 2 points $n=11$ et $k=2$ et pour la LSS à 3 points $n=11$ et $k=3$ ce qui donne respectivement des valeurs N_1 , N_2 , N_3 égale à 11, 55 et 165 jeux de données ce qui revient à construire au total 231 jeux de données.

Les valeurs manquantes présentes dans les jeux de données sont dues aux suppressions des valeurs BLOQ mais aussi à des valeurs non récoltées par la technique FRET. La méthode d'imputation par la médiane a été utilisée pour traiter ces valeurs manquantes car elle est plus robuste aux valeurs extrêmes qui peuvent être présentes dans les données. Les données ont ensuite été mises à l'échelle à l'aide de la méthode de standardisation. Enfin, une ACP a été réalisée sur les variables afin de créer différentes composantes principales et qui a permis de diminuer le nombre de variables à fournir au modèle. Cette ACP a également permis d'éliminer la multicolinéarité entre ces composantes. Les composantes les plus importantes ont ensuite été sélectionnées en utilisant la *Recursive Feature Elimination* (RFE, une technique de sélection de variables basée sur le principe de la sélection descendante) couplée à la validation croisée *Leave One Out* (RFECV-LOOCV) qui a permis par la même occasion de construire les modèles de régression linéaire multiple.

La RFE⁷² est une technique statistique qui sélectionne les variables selon leur importance à chaque itération. Dans le cadre de la régression linéaire multiple, cette sélection se fait sur la valeur absolue des coefficients des variables explicatives (des concentrations étudiées dans le cadre de cette étude). Une valeur absolue faible d'un coefficient indique qu'elle est peu informative sur la variable à prédire.

Une validation croisée *leave-one-out* est effectuée en même temps que la RFE. Cette réalisation simultanée de ces deux techniques permet de calculer la performance moyenne du modèle. Cette performance est alors égale à la

performance du modèle obtenue en validation croisée à la LOOCV du modèle ayant les meilleures variables sélectionnées par la méthode RFE (Tableau 21).

À l'issue de cette première sélection de variables, une suppression des variables non-significatives a été effectuée. Cette suppression s'est révélée importante car elle permet d'obtenir un modèle avec des variables interprétables et dont l'intervalle de confiance de leurs valeurs ne contient pas la valeur 0. La suppression des variables a été faite de manière itérative avec une *p-value* inférieure à 5% comme critère d'évaluation (à chaque itération la *p-value* la plus grande et supérieure à 5% est retirée jusqu'à n'obtenir que des variables ayant des *p-value* < 5%).

De plus, les intervalles de confiance à 95% des coefficients des composantes et de l'intercept ont été calculés afin d'évaluer la variabilité de ces coefficients. Les intervalles de confiance des coefficients à chaque itération de la LOOCV des meilleures composantes ont été calculés.

Au final, l'intervalle de confiance conservateur de chaque coefficient et de l'intercept est représenté par :

- la valeur minimale des bornes inférieures des intervalles de confiance des coefficients et de l'intercept toutes itérations confondues pour la borne inférieure.

- la valeur maximale des bornes supérieures des intervalles de confiance des coefficients et de l'intercept toutes itérations confondues pour la borne supérieure.

Tableau 21 : Tableau illustrant la RFECV-LOOCV.

Modèles	VARIABLES sélectionnées selon les performances moyennes en validation croisée <i>leave-one-out</i>
Modèle 1	X1, X2, X3
Modèle 2	X1, X2
Modèle 3	X2

L'évaluation du modèle a été réalisée à l'aide des métriques suivantes : r^2 , $r^2_{\text{Ajusté}}$, de la $RMSE\%$, de la $MPE\%$ et de la métrique J. La lecture de la littérature scientifique sur la LSS a permis d'identifier des seuils sur ces différentes métriques (Tableau 22) :

Tableau 22 : Tableau de référencement des données de la littérature sur le choix des seuils des métriques d'évaluation des modèles de LSS.

Substance	r^2	RMSE (Précision)	MPE (Biais)	Impact Factor
Mycophenolic Acid ⁷³	0,75<	<15%	<15%	6,251
Tacrolimus ⁵³	X	X	<15%	3,4
Tuberculosis drugs and moxifloxacin ⁷⁴	Le + haut (0,95)	<15%	<15%	10,8
Mycophenolic Acid ⁷⁵	0,901	X	X	6,1
Sirolimus ⁷⁶	X	<15%	<15%	9,3
IV Busulfan ⁶⁶	0,9	<15%	<15%	4,3
E7070 ⁶⁵	X	<15%	X	3,4
Factor IX ⁷⁷	X	<25%	<5%	10,4
Saroglitazar ⁶⁷	0,9	<30%	<30%	3,2
Levofloxacin ⁵¹	0,9	X	X	4,9

Au regard de ce tableau, les modèles ayant un r^2 supérieur à 0.9, une RMSE inférieure à 15%, une MPE inférieure à 15% et une métrique J la plus proche de 0 seront considérés comme de bons modèles. Le modèle retenu sera ensuite validé en vérifiant les hypothèses de linéarité, d'homoscédasticité des résidus, de normalité des résidus, de non-autocorrélation des résidus, de non multi-colinéarité et en analysant les graphiques de Bland-Altman et le graphique des *predicted vs observed*.

4.3. Résultats

4.3.1. Analyse descriptive des données

4.3.1.1. Analyse exploratoire

Une analyse exploratoire des données a été réalisée grâce aux statistiques descriptives en calculant l'effectif, la moyenne, l'écart-type, les valeurs minimales et maximales ainsi que les différents quartiles. Le Tableau 23 décrit les données :

Tableau 23 : Tableau descriptif des statistiques des variables explicatives.

	Effectif	Moyenne	Ecart type	min	25%	50%	75%	max
AUC (T _p .min/mL)	34	16907	7513,132	6633	11853,64	15678,84	20308,87	40442,94
C0 (T _p /mL)	34	46,35882	12,33998	25,95	36,84	46,975	55,1575	70,13
C5 (T _p /mL)	34	45,55199	11,64025	25,95127	36,84048	46,12759	52,17575	67,91332
C30 (T _p /mL)	34	45,02191	9,786253	28,52892	37,82043	43,06912	52,88226	59,99482
C60 (T _p /mL)	32	44,10765	10,26949	25,50109	36,16309	43,67558	52,64735	67,04703
C90 (T _p /mL)	33	40,11254	11,86235	22,39284	30,44897	38,56866	49,29007	69,69125
C120 (T _p /mL)	30	36,25352	10,39751	18,27185	28,96104	38,09256	45,1106	58,87362
C180 (T _p /mL)	31	32,79558	9,69708	16,82095	25,24069	32,58981	37,63182	58,74301
C240 (T _p /mL)	33	22,89566	6,238611	14,94469	16,74836	22,79864	28,49267	34,197
C360 (T _p /mL)	30	18,13921	4,595008	10,00878	15,6	17,1776	20,47016	27,18018
C600 (T _p /mL)	21	13,68212	3,997449	6,492053	10,60879	15,6	16,7	19,44568
C1440 (T _p /mL)	34	12,98824	4,161422	6	7,6	15,6	16,6	16,7
Poids (g)	34	471,6471	67,53266	390	423,25	452	516,5	673
Dose (T _p)	34	943,2941	135,0653	780	846,5	904	1033	1346
Âge (semaines)	34	10,82353	2,779477	8	9	9	12,75	18
Captex Dil (%w/w)	34	6,911765	1,519921	5,5	5,5	5,5	8,5	8,5
Captex Did (%w/w)	34	6,911765	1,519921	5,5	5,5	5,5	8,5	8,5
Kolliphore (%w/w)	34	10,46471	1,114609	9,3	9,3	11,5	11,5	11,5
Eau purifiée (%w/w)	34	19,51176	1,925234	17,5	17,5	21,3	21,3	21,3
PEG (%w/w)	34	0,211765	0,291043	0	0	0	0,6	0,6
SA (%w/w)	34	0,029412	0,04625	0	0	0	0,1	0,1
Taille (nm)	34	66,47059	17,73241	50	50	50	85	85
Potentiel zeta (mV)	34	8,947059	12,2529	-7,1	-1,3	4,2	18	30

4.3.1.2. Paramètres pharmacocinétiques

Les paramètres pharmacocinétiques par type et taille ont également été calculés à l'aide de PKanalix (lixoft 2021) et sont décrits dans le Tableau 24.

Tableau 24 : Tableau présentant les paramètres pharmacocinétiques des différentes NCL.

	50 NORM	50 PEG	50 SA	85 NORM	85 PEG	85 SA	Toutes
AUC _{Inf} (Tp.min/mL)	18005	8617	14355	15893	28371	15848	16 848
AUC _{Last} (Tp.min/mL)	9858	5290	9295	13434	16852	13687	11 403
Demi-vie (min)	287	109	216	184	388	192	229
MRT (min)	410	159	312	284	546	277	331
CL (mL/min)	0,06	0,12	0,073	0,058	0,031	0,062	0,07
V _d (mL)	22,19	17,7	21,84	14,64	15,39	17,15	18,15
Dose (Tp)	1000	982,33	999,33	891	827,33	968	944,67

Ces paramètres calculés montrent que les NCL 85-PEG ont l'exposition (AUC) la plus grande et à l'inverse que les NCL 50-PEG ont l'exposition la plus faible.

Concernant l'élimination quantifiée par la clairance, les NCL 50-PEG sont éliminées très rapidement tandis que les NCL 85-PEG sont éliminées plus lentement.

Le MRT le plus long a été observé pour les NCL 85 PEG tandis le MRT le plus court a été observé pour les NCL 50 PEG.

Concernant la distribution, les NCL 50 Norm se distinguent par leur meilleure distribution alors que les NCL 85 Norm affichent la distribution la plus faible.

Tous ces paramètres vont dans le même sens que les différents résultats obtenus dans l'étude de Lebreton *et al*³⁵.

4.3.1.3. Boîtes à moustaches

Les boîtes à moustaches des différentes variables ont été construites. Ces boîtes à moustaches ont peu de valeurs aberrantes. Les valeurs qui sont hors des queues ne sont pas réellement des valeurs aberrantes car elles

peuvent être observées en clinique. Les boîtes à moustaches des différentes variables sont dans l'Annexe 1.

4.3.1.4. Coefficient de corrélation de Bravais-Pearson et de Spearman

Les coefficients de corrélation de Bravais-Pearson et de Spearman à l'AUC ont été calculés pour chaque variable et ont permis de trouver les variables qui ont une relation linéaire ou non avec l'AUC (Tableau 25).

Tableau 25 : Tableau des coefficients de Bravais-Pearson et de Spearman des variables explicatives.

Variables explicatives	Coefficient de Bravais-Pearson	Coefficient de Spearman
C0 (T _p /mL)	0,07	0,06
C5 (T _p /mL)	0,1	0,08
C30 (T _p /mL)	0,31	0,29
C60 (T_p/mL)	0,59	0,64
C90 (T_p/mL)	0,58	0,69
C120 (T_p/mL)	0,61	0,7
C180 (T_p/mL)	0,56	0,69
C240 (T_p/mL)	0,75	0,87
C360 (T_p/mL)	0,79	0,83
C600 (T _p /mL)	0,25	0,14
C1440 (T _p /mL)	0,2	0,26
Poids (g)	-0,25	-0,29
Dose (T _p)	-0,25	-0,29
Âge (semaines)	-0,12	-0,05
Captex Dil (%w/w)	0,47	0,46
Captex Did (%w/w)	0,47	0,46
Kolliphore (%w/w)	-0,47	-0,46
Eau purifiée (%w/w)	-0,47	-0,46
PEG (%w/w)	0,16	0,01
SA (%w/w)	-0,17	-0,14
Taille (nm)	0,47	0,46
Potentiel zeta (mV)	-0,07	0,06

Les variables ayant un coefficient de corrélation de Bravais-Pearson et/ou de Spearman supérieur à 0,5 (en gras dans le Tableau 25) seront considérées comme des variables ayant une bonne corrélation avec la variable cible.

Dans cette étude, les variables qui ont le plus d'influence sur l'AUC sont exclusivement des concentrations et plus précisément les concentrations récoltées aux temps 60, 90, 120, 180, 240 et 360 minutes.

Étonnamment, les caractéristiques physico-chimiques et les paramètres individuels des rats n'ont pas démontré de fortes corrélations avec l'AUC.

4.3.2. Résultats LSS à 1 point

Le meilleur modèle de la LSS à 1 point utilise le temps de prélèvement à $t = 360$ minutes. Il a un r^2 égal à 0,682 (<0.9), un $r^2_{\text{ajusté}}$ égal à 0,597, une RMSE égale à 25,18% qui est supérieure à 15%, une MPE égale à 5,10% qui est bien inférieure à 15% et une métrique J égale à 1,024 plutôt éloignée de 0 (Tableau 26).

Ce modèle a été identifié comme le meilleur modèle de la LSS à 1 point selon la métrique J mais il ne pourra pas être sélectionné car il ne répond pas aux critères de sélection des métriques.

Tableau 26 : Tableau des résultats de la LSS à 1 point.

Variabes	r^2	$r^2_{\text{ajusté}}$	RMSE (%)	MPE (%)	J
C360 ln(C360) Poids Dose Âge Captex Dil Captex Did Kolliphor® Eau purifiée PEG SA Taille Potentiel zeta	0,682	0,597	25,18	5,10	1,024

4.3.3. Résultats LSS à 2 points

Le meilleur modèle de la LSS à 2 points utilise les temps de prélèvement à $t = 180$ min et à $t = 360$ min. Il a un r^2 égal à 0,647 (<0.9), un $r^2_{\text{ajusté}}$ égal à 0,583, une RMSE égale à 28,42% qui est supérieure à 15%, une MPE égale à 5,89% (<<15%) et une métrique J égale à 1,113 moins éloigné de 0 que la métrique J du meilleur modèle de la LSS à 1 point (Tableau 27).

Ce modèle a été identifié comme le meilleur modèle de la LSS à 2 points selon la métrique J mais il ne pourra pas être sélectionné car il ne répond pas aux critères de sélection des métriques.

Tableau 27 : Tableau des résultats de la LSS à 2 points.

Variabes	r^2	$r^2_{\text{ajusté}}$	RMSE (%)	MPE (%)	J
C180 C360 C180 x C360 ln(C180) ln(C360) ln(C180) x ln(C360) Poids Dose Âge Captex Dil Captex Did Kolliphor® Eau purifiée PEG SA Taille Potentiel zeta	0,647	0,583	28,42	5,89	1,113

4.3.4. Résultats LSS à 3 points

Le meilleur modèle de la LSS à 3 points utilise les temps de prélèvement à $t = 180$ min, $t = 360$ min et $t = 600$ min. Il a un r^2 égal à 0,952 qui est supérieur à 0,9, un $r^2_{\text{ajusté}}$ égal à 0,925, une RMSE égale à 9.81% qui est inférieure à 15%, une MPE égale à -0,03% qui est bien inférieure à 15% et une métrique J égale à 0.222 relativement proche de 0 (Tableau 28).

Ce modèle a été identifié comme le meilleur modèle de la LSS à 3 points selon la métrique J. De plus, ce modèle répond aux critères de sélection des métriques. Le choix du modèle toutes LSS confondues sera donc le meilleur modèle de la LSS à 3 points.

Ce modèle utilise 3 temps de prélèvement ($t = 180$ min, $t = 360$ min et $t = 600$ min) sachant que c'était le nombre maximum de points souhaité initialement.

Tableau 28 : Tableau des résultats de la LSS à 3 points.

Variabes	r^2	$r^2_{\text{ajusté}}$	RMSE (%)	MPE (%)	J
C180					
C360					
C600					
C180-C360					
C180-C600					
C360-C600					
$\ln(\text{C180}) \times \ln(\text{C360})$					
$\ln(\text{C180}) \times \ln(\text{C600})$					
$\ln(\text{C360}) \times \ln(\text{C600})$					
$\ln(\text{C180})$					
$\ln(\text{C360})$					
$\ln(\text{C600})$					
$\text{C180} \times \text{C360} \times \text{C600}$					
$\ln(\text{C180}) \times \ln(\text{C360}) \times \ln(\text{C600})$					
Poids	0,952	0,925	9,81	-0,03	0,222
Dose					
Âge					
Captex® Dil					
Captex® Did					
Kolliphor®					
Eau purifiée					
PEG					
SA					
Taille					
Potentiel zeta					

4.4. Validation du modèle choisi

4.4.1. Coefficients du modèle et composition des composantes principales sélectionnées

La RFE-LOOCV a permis de trouver les composantes principales qui donnent les meilleures performances des différents modèles développés. Ensuite, une élimination itérative des composantes non significatives a été effectuée.

Pour le modèle sélectionné, les composantes principales significatives retrouvées PC pour *Principal Component (Composantes principales)* sont au nombre de 12 à savoir : PC1, PC3, PC4, PC5, PC6, PC9, PC10, PC11, PC12, PC14, PC18, PC19.

Les compositions de chaque composante sont référencées en Annexe 3. Les valeurs des coefficients du modèle de régression linéaire multiple sur les composantes principales ainsi que leurs intervalles de confiance à 95% sont référencés dans l'Annexe 2.

4.4.2. Predicted vs observed

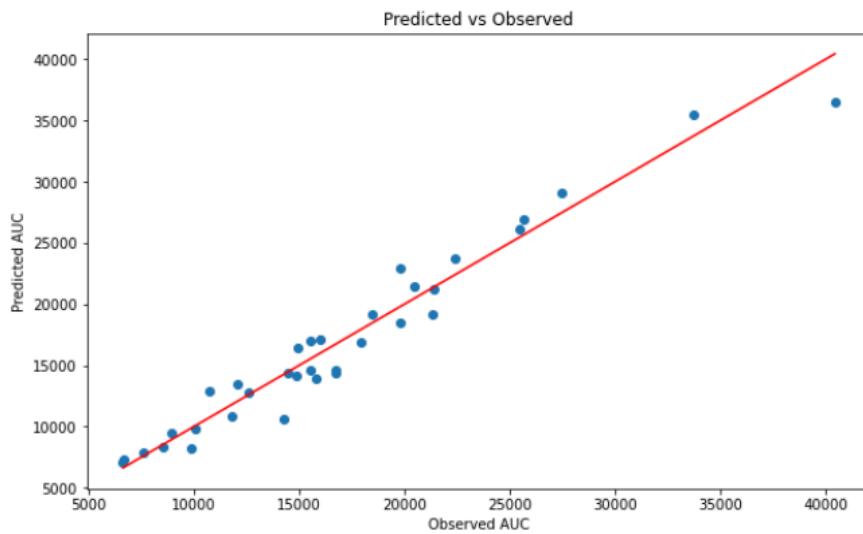


Figure 35 : Graphique *predicted vs observed*

Ce graphique (Figure 35) montre que le modèle a une bonne capacité de prédiction étant donné que les points sont proches de la droite d'identité $y = x$.

4.4.3. Vérification de la linéarité du modèle

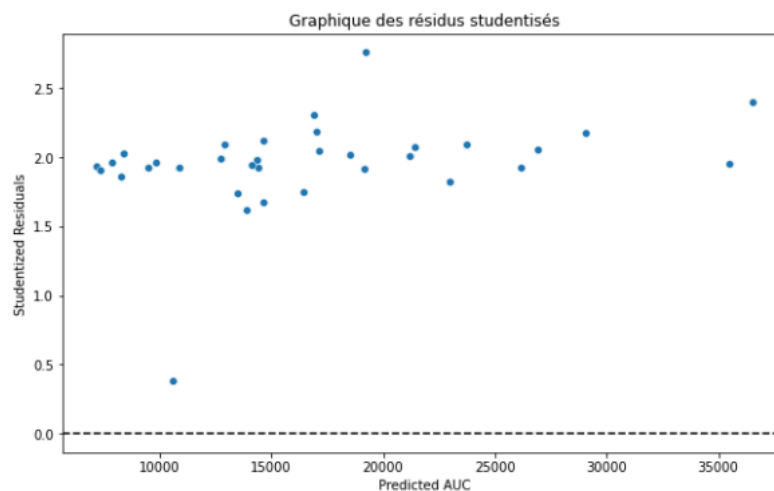


Figure 36 : Graphique des résidus studentisés.

Le graphique des résidus « *studentisés* » (Figure 36) nous montre que les résidus « *studentisés* » sont compris entre 0 et 3. Cet intervalle est bien compris dans l'intervalle $[-3 ; 3]$ ce qui confirme la linéarité du modèle.

4.4.4. Vérification de la normalité des résidus

Un Test de Shapiro-Wilk sur les résidus a été effectué. Le résultat de la statistique du test est égal à 0,97 avec une *p-value* égale à 0,38. Cette *p-value* est supérieur à 5% ce qui indique qu'on ne peut pas rejeter l'hypothèse nulle qui indique que la distribution des résidus suit une loi

normale. L'hypothèse de normalité des résidus est donc vérifiée au risque $\alpha=5\%$.

4.4.5. Vérification de l'homoscédasticité des résidus

Un Test de Breusch-Pagan sur les résidus a été effectué. Le résultat de la statistique du test est égal à 16,68 avec une *p-value* égale à 0,16. Cette *p-value* est supérieure à 5% ce qui permet de ne pas rejeter l'hypothèse nulle H_0 d'homoscédasticité des résidus.

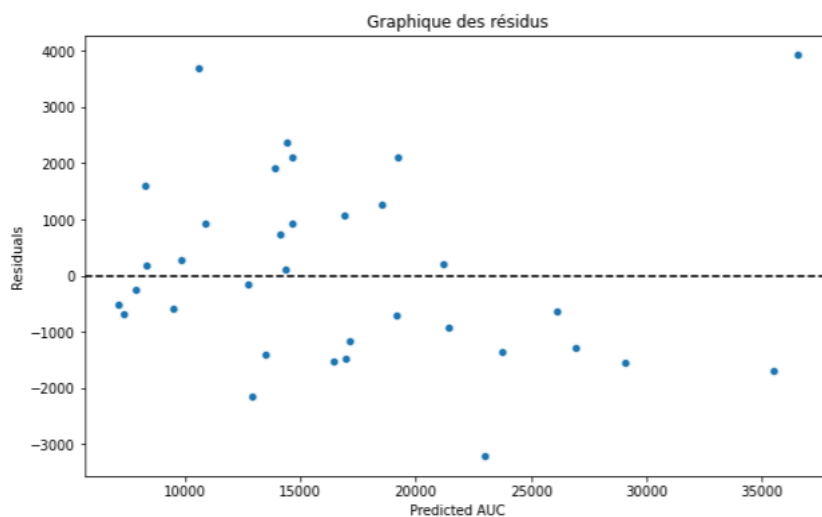


Figure 37 : Graphique des résidus.

De plus, le graphique des résidus du modèle (Figure 37) ne montre pas de schéma ayant une forme particulière.

L'hypothèse d'homoscédasticité des résidus est donc acceptée au risque $\alpha=5\%$.

4.4.6. Vérification de la non-autocorrélation des résidus

Un Test de Durbin-Watson sur les résidus a été effectué. Le résultat de la statistique du test est égal à 1,64. La valeur est proche de 2 ce qui valide l'hypothèse qu'il n'y pas d'autocorrélation significative dans les résidus. La non-autocorrélation des résidus du modèle est donc confirmée.

4.4.7. Vérification de la non multi-colinéarité

La vérification de la non multi-colinéarité a été réalisé en calculant les VIFs. Pour cela une LOOCV a été effectuée, les VIFs ont été calculés pour chaque composante à chaque itération. La valeur maximale des VIFs de chaque itération de la LOOCV toutes composantes prises en compte est de 1.46 ce qui est inférieur à 5 (Tableau 29). Il n'y a donc pas de multi-colinéarité entre les différentes composantes principales sélectionnées par la méthode RFE-LOOCV. Cette absence de multi-colinéarité est due principalement à l'ACP qui a permis d'obtenir des combinaisons linéaires des variables initiales.

Tableau 29 : VIF maximal des variables toutes itérations confondues

Variables	VIF maximal
PC1	1,46
PC3	1,15
PC4	1,07
PC5	1,09
PC6	1,08
PC9	1,24
PC10	1,13
PC11	1,15
PC12	1,21
PC14	1,18
PC18	1,11
PC19	1,19

4.4.8. Bland-Altman

Un Bland-Altman (Figure 38) a été tracé afin de comparer la technique de référence de calcul de l'AUC à la méthode de calcul de l'AUC développée en utilisant le modèle à 3 points sélectionné. La technique de référence calcule l'AUC avec le logiciel PKanalix (lixoft 2021) qui utilise toutes les concentrations récoltées.

Sur ce graphique, le biais tracé est très proche de 0 ce qui indique qu'il y a un biais faible entre les deux méthodes de calcul de l'AUC. On remarque également que la plupart des points sont compris dans l'intervalle de confiance ce qui indique que les deux méthodes convergent vers des résultats semblables.

Au vu de ces résultats, les deux techniques de calcul peuvent être utilisées pour calculer les AUCs des différents types de NCL.

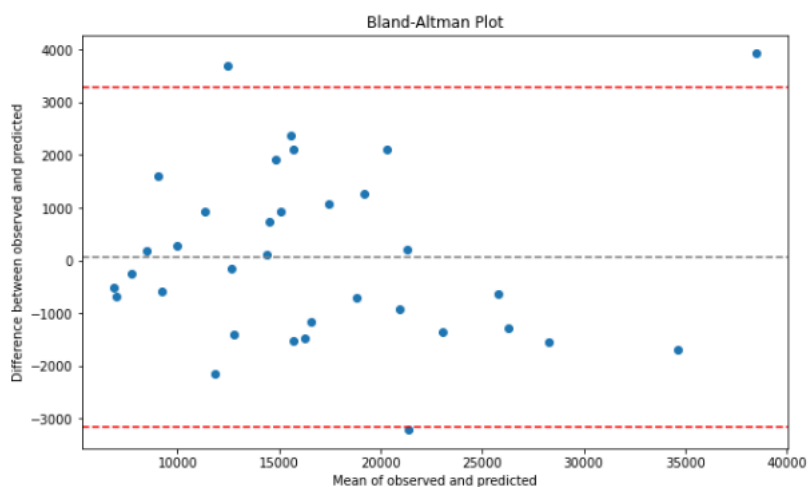


Figure 38 : Tracé de Bland-Altman pour la concordance entre les AUCs mesurées et les AUCs prédites par le modèle de LSS à 3 points.

4.4.9. Significativité du modèle

La significativité du modèle a été calculée à chaque itération. La valeur maximale de la p-value du modèle toutes itérations prises en compte est de $1,93 \times 10^{-14}$. Cette valeur maximale est inférieure à 5% ce qui indique que les différents modèles sont significatifs à chaque itération.

4.5. Discussion

Le modèle final choisi témoigne de la complexité de la pharmacocinétique des NCL car il prend en compte de nombreux facteurs à savoir les concentrations plasmatiques, les caractéristiques physico-chimiques des NCLs et les caractéristiques individuelles des rats.

En plus de la complexité des variables, le modèle obtenu est d'autant plus complexe car une ACP a été réalisée afin de trouver un modèle avec un faible biais et une bonne précision. Ce modèle confirme que la pharmacocinétique des NCL dépend de leur type et de leur taille, comme l'a démontré l'étude de Lebreton *et al*³⁵.

La construction basée sur les composantes principales a permis de trouver, pour la première fois, un modèle qui répond aux critères de sélection. Cependant les composantes principales sont difficiles à interpréter car ce sont des combinaisons linéaires des variables initiales. En effet, l'effet de l'augmentation d'une des variables initiales sur l'AUC est difficilement calculable car chaque variable est incluse dans les composantes qui ont elles-mêmes des coefficients (poids) différents. Par exemple, l'augmentation du potentiel zêta impactera toutes les composantes qui elles-mêmes ont des coefficients différents entre elles, l'interprétation en devient difficile. Pour avoir une idée de l'impact de l'augmentation d'une des variables sur l'AUC, il faut se pencher sur les coefficients de corrélation de Bravais-Pearson et de Spearman qui donnent une bonne indication de la relation entre une variable explicative et l'AUC. Par exemple l'augmentation du potentiel zêta n'induit pas d'effet significatif sur l'AUC car la valeur absolue de ces coefficients de corrélation de Bravais-Pearson et de Spearman sont inférieurs à 5 respectivement de -0.07 et 0.06.

Le modèle trouvé est donc dépendant de 3 temps de prélèvements à savoir : 180, 360 et 600 minutes. En regardant de plus près, ces temps sont très proches du $t_{1/2}$ moyen des NCL tous types et toutes tailles confondus qui est de 229 minutes. Ce qui signifie que ces temps correspondent plutôt à la phase d'élimination car environ 50% des NCL ont été éliminées à ces temps. Ce constat est également soutenu par le fait que dans une administration IV la phase d'absorption est souvent négligée comme c'est le cas dans cette étude. Ainsi, pour mieux comprendre et optimiser l'exposition des NCL, il faut d'avantage s'attarder sur la phase d'élimination et donc se pencher sur la compréhension de l'impact de leur instabilité dans les fluides biologiques. Il est également important de comprendre plus en profondeur et si possible de mettre en évidence les mécanismes d'élimination hépatobiliaire et les mécanismes d'élimination rénale des NCL.

De plus, les temps sélectionnés encadrent le MRT moyen qui est égal à 331 minutes. Pour rappel, le MRT représente la durée de l'ensemble des phases pharmacocinétiques. Ce résultat est cohérent avec le fait que ces temps soient considérés comme les plus informatifs.

Le modèle trouvé pourra être utilisé pour déterminer l'exposition (AUC) de nouvelles NCL. Il faudra pour cela récolter les concentrations aux temps 180, 360 et 600 minutes, les caractéristiques physico-chimiques des NCL ainsi que les caractéristiques individuelles des rats utilisés pour pouvoir calculer l'AUC.

Ce travail peut faire l'objet de plusieurs améliorations. Premièrement une augmentation du nombre de données pourrait conduire à de meilleurs résultats, car le nombre de données de cette étude est faible. Ainsi l'augmentation du nombre de données pourrait permettre au modèle de mieux capter les relations et affiner le calcul des coefficients des variables du modèle de LSS. La simulation de données à l'aide du logiciel Simulix (Lixoft 2021, Paris, France) pourrait être envisagée afin d'obtenir plus de données. Cependant, bien que ces nouvelles données obtenues par simulation puissent améliorer les performances du modèle, elles présentent aussi un risque non négligeable de générer un biais dans l'analyse.

De plus, la méthode d'imputation pourrait être optimisée. En effet, de nombreuses méthodes d'imputation plus avancées existent comme décrit brièvement dans la partie « Traitement des valeurs manquantes » du manuscrit. Ces méthodes pourraient être testées afin d'améliorer potentiellement les résultats du modèle. Elles pourraient également être envisagées pour remplacer les BLOQs et les LLOQs ce qui éviterait de supprimer les BLOQs afin de conserver un maximum d'information.

Une autre perspective serait de créer de nouvelles variables qui se rapprochent du calcul de l'AUC selon l'approche log trapézoïdal. En effet, cette approche utilise une équation qui dépend des concentrations logarithmiques mais aussi des concentrations non logarithmiques⁷⁸. Pour la création de ces variables, notons $f(C_i, C_{i+1})$ la fonction qui calcule l'aire sous la courbe entre deux points selon l'approche log trapézoïdal. Une nouvelle variable pourrait être par exemple les résultats du calcul $f(C_{120}, C_{180})$ qui représenteraient les AUCs entre les points 120 minutes et 180 minutes. De telles variables pourraient potentiellement donner de meilleurs résultats car elles se rapprocheraient du calcul de l'AUC utilisé.

Une autre idée de modélisation serait de réaliser une LSS selon l'approche bayésienne. Cette approche est différente de la régression linéaire classique et permettrait de comparer les deux approches et de choisir le meilleur des deux modèles.

Enfin, la métrique J pourrait être améliorée et intégrer de nouveaux éléments tels que :

- Les valeurs des résidus *studentisés*
- Une mesure de forme particulière du graphique des résidus
- Les résultats des tests de normalité, d'homoscédasticité des résidus et de non-autocorrélation des résidus
- Les *Variance Inflation Factor* (VIFs)
- Les *p-values* des variables et du modèle
- Le nombre de points à l'extérieur des intervalles de confiance du Bland-Altman
- Le biais calculé dans le Bland-Altman
- Tout autre métriques ou moyen quantifiable pour valider un modèle de régression linéaire multiple.

Des perspectives médicales sont également envisageables. Le modèle obtenu peut permettre de faire du suivi thérapeutique. En effet, les 3 temps de prélèvements permettent d'établir le profil pharmacocinétique des NCL avec une bonne précision et un faible biais. Le suivi thérapeutique pourra donc se faire à l'aide de ces 3 temps à savoir 180, 360 et 600 minutes. Ce modèle permet également de faire de l'adaptation de dose. En effet, la dose est une variable explicative du modèle trouvé et permet donc de réguler l'exposition des NCL.

Des perspectives en termes de formulation peuvent aussi être envisagées. Il existe des nanoparticules appelées particules Janus qui sont des nanoparticules dont la surface est biphasique (par exemple des nanoparticules PEG+SA). Il serait intéressant de tester ce modèle afin de voir s'il permet aussi d'expliquer la pharmacocinétique de ce type de nanoparticules.

De plus, ce modèle pourrait aussi être utilisé pour évaluer l'exposition des NCL de composition différente en PEG/SA ou encore de taille différente.

Enfin, il serait intéressant de réaliser une LSS pour des NCL contenant un principe actif (nanomédicament) ce qui permettrait de comparer les temps de prélèvements sélectionnés avec ceux de la LSS du principe actif seul. Cette approche permettrait donc d'évaluer l'impact de la nanocapsule sur la pharmacocinétique du principe actif étudié.

La LSS est donc une technique statistique très puissante permettant de réduire le nombre de temps de prélèvements tout en gardant une bonne précision et un faible biais. Cette technique peut être utilisée dans de nombreux domaines médicaux.

Conclusion

En résumé, après avoir défini la notion de nanomédicament, ce manuscrit a apporté des éléments de compréhension sur la pharmacocinétique de ces derniers qui sont indispensables afin de réaliser une LSS. S'en est suivi une revue des différentes étapes et approches pour effectuer la LSS. Enfin, une étude visant à développer une LSS sur des NCL a été abordée. Au cours de ce travail, un modèle de régression linéaire multiple sur composantes principales couplé à une RFECV-LOOCV a été développé pour la première fois, dont les variables explicatives sont des combinaisons linéaires de variables dépendantes des concentrations plasmatiques, des paramètres physico-chimiques des NCL mais aussi des caractéristiques individuelles des rats.

Le modèle trouvé dépend de 3 temps de prélèvements à savoir les temps 180, 360 et 600 minutes. Ce modèle a de bonnes performances avec un r^2 égal à 0.952 qui est supérieur à 0.9, une RMSE égale à 9.81 qui est inférieure à 15%, une MPE égale à -0.03 qui est inférieure à 15%. Ce modèle a également un $r^2_{\text{ajusté}}$ égal à 0.925 et une métrique J égale à 0.222 qui est proche de 0. Il a été validé à l'aide de différents outils de validations statistiques et de validation graphiques qui ont démontré que le modèle développé respecte les hypothèses de la régression linéaire multiple. Ce modèle a également démontré qu'il peut servir de technique de mesure avec des performances comparables à la technique de calcul de l'AUC de PKanalix utilisant tous les temps de prélèvements.

Les deux conclusions principales de ce travail résident d'une part dans l'assurance que la pharmacocinétique des nanomédicaments est complexe car elle dépend de nombreux paramètres et interactions. Et d'autre part, que celle-ci dépend essentiellement de la phase d'élimination qui devient la clé de compréhension de l'exposition des NCL lipidiques et doit faire l'objet d'études plus approfondies pour optimiser l'exposition aux NCL.

La modélisation qui est, elle, purement mathématique, est difficilement interprétable pour toute personne ayant pas ou peu de notion de pharmacocinétique mais aussi de galénique dans ce cas précis des nanomédicaments. En effet, Les variables explicatives du modèle tout comme l'interprétation du modèle lui-même sont difficilement interprétables. Le pharmacien ayant des connaissances pour interpréter ces modèles obtenus devient donc un acteur majeur pour la compréhension et l'utilisation de ces modèles. Ce travail montre donc l'importance du rôle du pharmacien dans l'étude de la pharmacocinétique des NCL.

Bibliographie

1. Lebreton V, Legeay S, Saulnier P, Lagarce F. Specificity of pharmacokinetic modeling of nanomedicines. *Drug Discovery Today*. 2021;26(10):2259-2268. doi:10.1016/j.drudis.2021.04.017
2. Fanciullino R, Gattacceca F, Rodallec A, Ciccolini J. Pharmacocinétique des nanomédicaments. In: *Pharmacocinétique: Les Fondamentaux*. Edimark santé. ; 2018:408.
3. Parodi A, Buzaeva P, Nigovora D, et al. Nanomedicine for increasing the oral bioavailability of cancer treatments. *J Nanobiotechnol*. 2021;19(1). doi:10.1186/s12951-021-01100-2
4. Xiao X, Teng F, Shi C, et al. Polymeric nanoparticles—Promising carriers for cancer therapy. *Front Bioeng Biotechnol*. 2022;10. doi:10.3389/fbioe.2022.1024143
5. Haripriyaa M, Suthindhiran K. Pharmacokinetics of nanoparticles: current knowledge, future directions and its implications in drug delivery. *Futur J Pharm Sci*. 2023;9(1):113. doi:10.1186/s43094-023-00569-y
6. Zenze M, Daniels A, Singh M. Dendrimers as Modifiers of Inorganic Nanoparticles for Therapeutic Delivery in Cancer. *Pharmaceutics*. 2023;15(2):398. doi:10.3390/pharmaceutics15020398
7. Wang J, Li B, Qiu L, Qiao X, Yang H. Dendrimer-based drug delivery systems: history, challenges, and latest developments. *J Biol Eng*. 2022;16(1):18. doi:10.1186/s13036-022-00298-5
8. Perumal S, Atchudan R, Lee W. A Review of Polymeric Micelles and Their Applications. *Polymers*. 2022;14(12):2510. doi:10.3390/polym14122510
9. Wang Q, Atluri K, Tiwari AK, Babu RJ. Exploring the Application of Micellar Drug Delivery Systems in Cancer Nanomedicine. *Pharmaceutics*. 2023;16(3):433. doi:10.3390/ph16030433
10. Liu Y, Castro Bravo KM, Liu J. Targeted liposomal drug delivery: a nanoscience and biophysical perspective. *Nanoscale Horiz*. 2021;6(2):78-94. doi:10.1039/D0NH00605J
11. Beltrán-Gracia E, López-Camacho A, Higuera-Ciapara I, Velázquez-Fernández JB, Vallejo-Cardona AA. Nanomedicine review: clinical developments in liposomal applications. *Cancer Nano*. 2019;10(1):11. doi:10.1186/s12645-019-0055-y
12. Deng S, Gigliobianco MR, Censi R, Di Martino P. Polymeric Nanocapsules as Nanotechnological Alternative for Drug Delivery System: Current Status, Challenges and Opportunities. *Nanomaterials*. 2020;10(5):847. doi:10.3390/nano10050847
13. Huynh NT, Passirani C, Saulnier P, Benoit JP. Lipid nanocapsules: A new platform for nanomedicine. *International Journal of Pharmaceutics*. 2009;379(2):201-209. doi:10.1016/j.ijpharm.2009.04.026
14. Wei Q, Li P, Yang T, et al. The promise and challenges of combination therapies with antibody-drug conjugates in solid tumors. *J Hematol Oncol*. 2024;17(1):1. doi:10.1186/s13045-023-01509-2
15. Cho H, Jeon SI, Ahn CH, Shim MK, Kim K. Emerging Albumin-Binding Anticancer Drugs for Tumor-Targeted Drug Delivery: Current Understandings and Clinical Translation. *Pharmaceutics*. 2022;14(4):728. doi:10.3390/pharmaceutics14040728
16. Xu M, Han X, Xiong H, et al. Cancer Nanomedicine: Emerging

- Strategies and Therapeutic Potentials. *Molecules*. 2023;28(13):5145. doi:10.3390/molecules28135145
17. Omidian H, Babanejad N, Cubeddu LX. Nanosystems in Cardiovascular Medicine: Advancements, Applications, and Future Perspectives. *Pharmaceutics*. 2023;15(7):1935. doi:10.3390/pharmaceutics15071935
 18. Ghorai S, Shand H, Patra S, et al. Nanomedicine for the Treatment of Viral Diseases: Smaller Solution to Bigger Problems. *Pharmaceutics*. 2024;16(3):407. doi:10.3390/pharmaceutics16030407
 19. Jagaran K, Singh M. Nanomedicine for Neurodegenerative Disorders: Focus on Alzheimer's and Parkinson's Diseases. *Int J Mol Sci*. 2021;22(16):9082. doi:10.3390/ijms22169082
 20. Yang Y, Santamaria P. Evolution of nanomedicines for the treatment of autoimmune disease: From vehicles for drug delivery to inducers of bystander immunoregulation. *Advanced Drug Delivery Reviews*. 2021;176:113898. doi:10.1016/j.addr.2021.113898
 21. Li M, Zou P, Tyner K, Lee S. Physiologically Based Pharmacokinetic (PBPK) Modeling of Pharmaceutical Nanoparticles. *AAPS J*. 2017;19(1):26-42. doi:10.1208/s12248-016-0010-3
 22. Sun R, Xiang J, Zhou Q, et al. The tumor EPR effect for cancer drug delivery: Current status, limitations, and alternatives. *Advanced Drug Delivery Reviews*. 2022;191:114614. doi:10.1016/j.addr.2022.114614
 23. Quéré I. Description anatomique et histologique, physiologie du système lymphatique. *La Presse Médicale*. 2010;39(12):1269-1278. doi:10.1016/j.lpm.2010.09.009
 24. F.Thomas, N.Picard, G.Ubeaud-Séquier. Métabolisme et élimination biliaire. In: *Pharmacocinétique: les fondamentaux*. Edimark santé; 2018:408.
 25. Padyana AK, Gross S, Jin L, et al. Structure and inhibition mechanism of the catalytic domain of human squalene epoxidase. *Nat Commun*. 2019;10(1):97. doi:10.1038/s41467-018-07928-x
 26. E.Chatelut, F.Thomas. Élimination rénale. In: *Pharmacocinétique: les fondamentaux*. Edimark santé; 2018:408.
 27. Jo DH, Kim JH, Lee TG, Kim JH. Size, surface charge, and shape determine therapeutic effects of nanoparticles on brain and retinal diseases. *Nanomedicine: Nanotechnology, Biology and Medicine*. 2015;11(7):1603-1611. doi:10.1016/j.nano.2015.04.015
 28. J.Ciccolini, G.Ubeaud-Séquier. Âge. In: *Pharmacocinétique: les fondamentaux*. Edimark santé; 2018:408.
 29. C.Lloret, X.Declèves. Morphologie et pharmacocinétique. In: *Pharmacocinétique: les fondamentaux*. Edimark santé; 2018:408.
 30. E.Chatelut. Paramètres pharmacocinétiques: définitions et méthodes de détermination. In: *Pharmacocinétique: les fondamentaux*. Edimark santé; 2018:408.
 31. S.Khier, F.Bourasset, F.Gattacceca. Paramètres graphiques après administration unique: approches compartimentale et non compartimentale. In: *Pharmacocinétique: les fondamentaux*. Edimark santé; 2018:408.
 32. E.Chatelut, F.Thomas, E.Curis. Administration unique et paramètres fondamentaux (primaires). In: *Pharmacocinétique: les fondamentaux*.

Edimark santé; 2018:408.

33. F.Bourasset, X.Declèves, E.Chatelut. Linéarité, stationnarité et pharmacocinétique non linéaire. In: *Pharmacocinétique : les fondamentaux*. Edimark santé; 2018:408.

34. S.Khier, E.Chatelut. Pharmacocinétique de population : principes généraux. In: *Pharmacocinétique : les fondamentaux*. Edimark santé; 2018:408.

35. Lebreton V, Kaeokhamloed N, Vasylaki A, et al. Pharmacokinetics of intact lipid nanocapsules using new quantitative FRET technique. *Journal of Controlled Release*. 2022;351:681-691. doi:10.1016/j.jconrel.2022.09.057

36. Roger E, Gimel JC, Bensley C, Klymchenko AS, Benoit JP. Lipid nanocapsules maintain full integrity after crossing a human intestinal epithelium model. *Journal of Controlled Release*. 2017;253:11-18. doi:10.1016/j.jconrel.2017.03.005

37. van den Elsen SHJ, Sturkenboom MGG, Akkerman OW, et al. Limited Sampling Strategies Using Linear Regression and the Bayesian Approach for Therapeutic Drug Monitoring of Moxifloxacin in Tuberculosis Patients. *Antimicrobial Agents and Chemotherapy*. 2019;63(7):10.1128/aac.00384-19. doi:10.1128/aac.00384-19

38. SAPORTA G. Description unidimensionnelle de données numériques. In: *Probabilités, analyse des données et Statistique*. TECHNIP; 2011:622.

39. Grégory Denglos. Statistique descriptive. In: *Statistiques et probabilités appliquées*. Quadrige. PUF; 2016:288.

40. Kieran Healy. *Data Visualization a practical introduction*. Princeton University Press; 2019.

41. 6.4. Imputation of missing values. scikit-learn. Accessed April 10, 2024. <https://scikit-learn/stable/modules/impute.html>

42. StandardScaler. scikit-learn. Accessed July 23, 2024. <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

43. MinMaxScaler. scikit-learn. Accessed July 23, 2024. <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>

44. Box GEP, Cox DR. An Analysis of Transformations.

45. power_transform. scikit-learn. Accessed July 23, 2024. https://scikit-learn/stable/modules/generated/sklearn.preprocessing.power_transform.html

46. SAPORTA G. L'analyse en composantes principales. In: *Probabilités, analyse des données et Statistique*. TECHNIP; 2011:622.

47. SAPORTA G. Description bidimensionnelle et mesures de liaison entre variables. In: *Probabilités, analyse des données et Statistique*. TECHNIP; 2011:622.

48. SAPORTA G. La régression linéaire multiple et le modèle linéaire général. In: *Probabilités, analyse des données et Statistique*. TECHNIP; 2011:622.

49. PIERRE-ANDRÉ CORNILLON, ERIC MATZNER-LOBER. La régression linéaire multiple. In: *Régression Théorie et applications*. Springer; :302.

50. La pharmacocinétique de population. PhinC Development. Accessed April 9, 2024. <https://www.phinc-development.com/fr/services/la-pharmacocinetique-de-population/>

51. Alsultan A, An G, Peloquin CA. Limited Sampling Strategy and Target

- Attainment Analysis for Levofloxacin in Patients with Tuberculosis. *Antimicrob Agents Chemother.* 2015;59(7):3800-3807. doi:10.1128/AAC.00341-15
52. L'apprentissage statistique : pourquoi, comment? In: *Apprentissage statistique*. Algorithmes. Eyrolles; 2008.
53. Mathew BS, Fleming DH, Jeyaseelan V, et al. A limited sampling strategy for tacrolimus in renal transplant patients. *Brit J Clinical Pharma.* 2008;66(4):467-472. doi:10.1111/j.1365-2125.2008.03251.x
54. Anne-Sophie Julien. Validation de modèles paramétriques. Atelier de formation presented at: CHU de Québec.
55. PIERRE-ANDRÉ CORNILLON, ERIC MATZNER-LOBER. Validation du modèle. In: *Régression Théorie et applications*. Springer; :302.
56. Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests.
57. Chen Y. Spatial Autocorrelation Approaches to Testing Residuals from Least Squares Regression. Schumann GJP, ed. *PLoS ONE*. 2016;11(1):e0146865. doi:10.1371/journal.pone.0146865
58. Van Boekel GAJ, Donders ART, Hoogtanders KEJ, Havenith TRA, Hilbrands LB, Aarnoutse RE. Limited sampling strategy for prolonged-release tacrolimus in renal transplant patients by use of the dried blood spot technique. *Eur J Clin Pharmacol.* 2015;71(7):811-816. doi:10.1007/s00228-015-1863-6
59. Tranmer M, Murphy J, Elliot M, Pampaka M. Multiple Linear Regression (2nd Edition).
60. Standard error using the Fisher Information Matrix. Monolix. Accessed July 24, 2024. <https://monolix.lixoft.com/tasks/standard-error-using-the-fisher-information-matrix/>
61. SAPORTA G. Méthodes de Monte-Carlo et de rééchantillonnage (Jack-knife, bootstrap). In: *Probabilités, analyse des données et Statistique*. TECHNIP; 2011:622.
62. Bergstrand M, Hooker AC, Wallin JE, Karlsson MO. Prediction-Corrected Visual Predictive Checks for Diagnosing Nonlinear Mixed-Effects Models. *AAPS J.* 2011;13(2):143-151. doi:10.1208/s12248-011-9255-z
63. Gaspari F, Anedda MF, Signorini O, Remuzzi G, Perico N. Prediction of cyclosporine area under the curve using a three-point sampling strategy after Neoral administration. *Journal of the American Society of Nephrology.* 1997;8(4):647-652. doi:10.1681/ASN.V84647
64. Wang P, Zhang Q, Zhu Z, et al. Population Pharmacokinetics and Limited Sampling Strategy for Therapeutic Drug Monitoring of Polymyxin B in Chinese Patients With Multidrug-Resistant Gram-Negative Bacterial Infections. *Front Pharmacol.* 2020;11:829. doi:10.3389/fphar.2020.00829
65. Van Kesteren C, Mathôt RAA, Raymond E, et al. Development and validation of limited sampling strategies for prediction of the systemic exposure to the novel anticancer agent E7070 (N-(3-chloro-7-indolyl)-1,4-benzenedisulphonamide): *Limited sampling strategies for E7070*. *British Journal of Clinical Pharmacology.* 2002;54(5):463-471. doi:10.1046/j.1365-2125.2002.01684.x
66. Dupuis LL, Sibbald C, Schechter T, et al. IV Busulfan Dose Individualization in Children undergoing Hematopoietic Stem Cell Transplant: Limited Sampling Strategies. *Biology of Blood and Marrow*

- Transplantation*. 2008;14(5):576-582. doi:10.1016/j.bbmt.2008.03.002
67. Joshi SN, Srinivas NR, Parmar DV. Limited Sampling Strategy for Accurate Prediction of Pharmacokinetics of Saroglitazar: A 3-point Linear Regression Model Development and Successful Prediction of Human Exposure. *Clinical Therapeutics*. 2018;40(3):456-468.e1. doi:10.1016/j.clinthera.2018.01.016
68. McEneny-King A, Chelle P, Goggans MH, et al. Limited sampling strategies for accurate determination of extended half-life factor VIII pharmacokinetics in severe haemophilia A patients. *Haemophilia*. 2021;27(3):408-416. doi:10.1111/hae.14288
69. Sobiak J, Resztak M, Chrzanowska M, Zachwieja J, Ostalska-Nowicka D. The Evaluation of Multiple Linear Regression-Based Limited Sampling Strategies for Mycophenolic Acid in Children with Nephrotic Syndrome. *Molecules*. 2021;26(12):3723. doi:10.3390/molecules26123723
70. Mizaki T, Nobata H, Banno S, et al. Population pharmacokinetics and limited sampling strategy for therapeutic drug monitoring of mycophenolate mofetil in Japanese patients with lupus nephritis. *J Pharm Health Care Sci*. 2023;9(1):1. doi:10.1186/s40780-022-00271-w
71. Kaeokhamloed N, Legeay S, Roger E. FRET as the tool for in vivo nanomedicine tracking. *Journal of Controlled Release*. 2022;349:156-173. doi:10.1016/j.jconrel.2022.06.048
72. RFE. scikit-learn. Accessed July 24, 2024. https://scikit-learn/stable/modules/generated/sklearn.feature_selection.RFE.html
73. Ting LSL, Partovi N, Levy RD, Riggs KW, Ensom MHH. Limited Sampling Strategy for Predicting Area Under the Concentration-Time Curve of Mycophenolic Acid in Adult Lung Transplant Recipients. *Pharmacotherapy*. 2006;26(9):1232-1240. doi:10.1592/phco.26.9.1232
74. Magis-Escurra C, Later-Nijland HMJ, Alffenaar JWC, et al. Population pharmacokinetics and limited sampling strategy for first-line tuberculosis drugs and moxifloxacin. *International Journal of Antimicrobial Agents*. 2014;44(3):229-234. doi:10.1016/j.ijantimicag.2014.04.019
75. Capone D, Tarantino G, Kadilli I, et al. Evaluation of mycophenolic acid systemic exposure by limited sampling strategy in kidney transplant recipients receiving enteric-coated mycophenolate sodium (EC-MPS) and cyclosporine. *Nephrology Dialysis Transplantation*. 2011;26(9):3019-3025. doi:10.1093/ndt/gfq819
76. Kaplan B, Meier-Kriesche HU, Napoli K, Kahan BD. A Limited Sampling Strategy for Estimating Sirolimus Area-Under-the-Concentration Curve. *Clinical Chemistry*. 1997;43(3):539-540. doi:10.1093/clinchem/43.3.539
77. Preijers T, Van Spengler MWF, Meijer K, et al. In silico evaluation of limited sampling strategies for individualized dosing of extended half-life factor IX concentrates in hemophilia B patients. *Eur J Clin Pharmacol*. 2022;78(2):237-249. doi:10.1007/s00228-021-03173-2
78. Data processing and calculation rules. PKANALIX. Accessed July 23, 2024. <https://pkanalix.lixoft.com/calculation-rules/>

Annexes

Annexe 1 : Box Plots des variables

Annexe 2 : Tableau des coefficients des composantes principales ainsi que leurs intervalles de confiance à 95%.

Annexe 3 : Composition des composantes principales sélectionnées pour le modèle de la LSS à 3 points

Annexe 4 : Composition du jeu de données initial

Annexe 5 : Code pour la description statistique des données

Annexe 6 : Code pour la création des box plots

Annexe 7 : Code pour le calcul des coefficients de corrélations

Annexe 8 : Code LSS 1 point

Annexe 9 : Code pour la création des datasets des meilleures composantes principales de la LSS 1 point

Annexe 10 : Code pour l'amélioration du meilleur modèle de la LSS 1 point

Annexe 11 : Code LSS 2 points

Annexe 12 : Code pour la création des datasets des meilleures composantes principales de la LSS 2 points

Annexe 13 : Code pour l'amélioration du meilleur modèle de la LSS 2 points

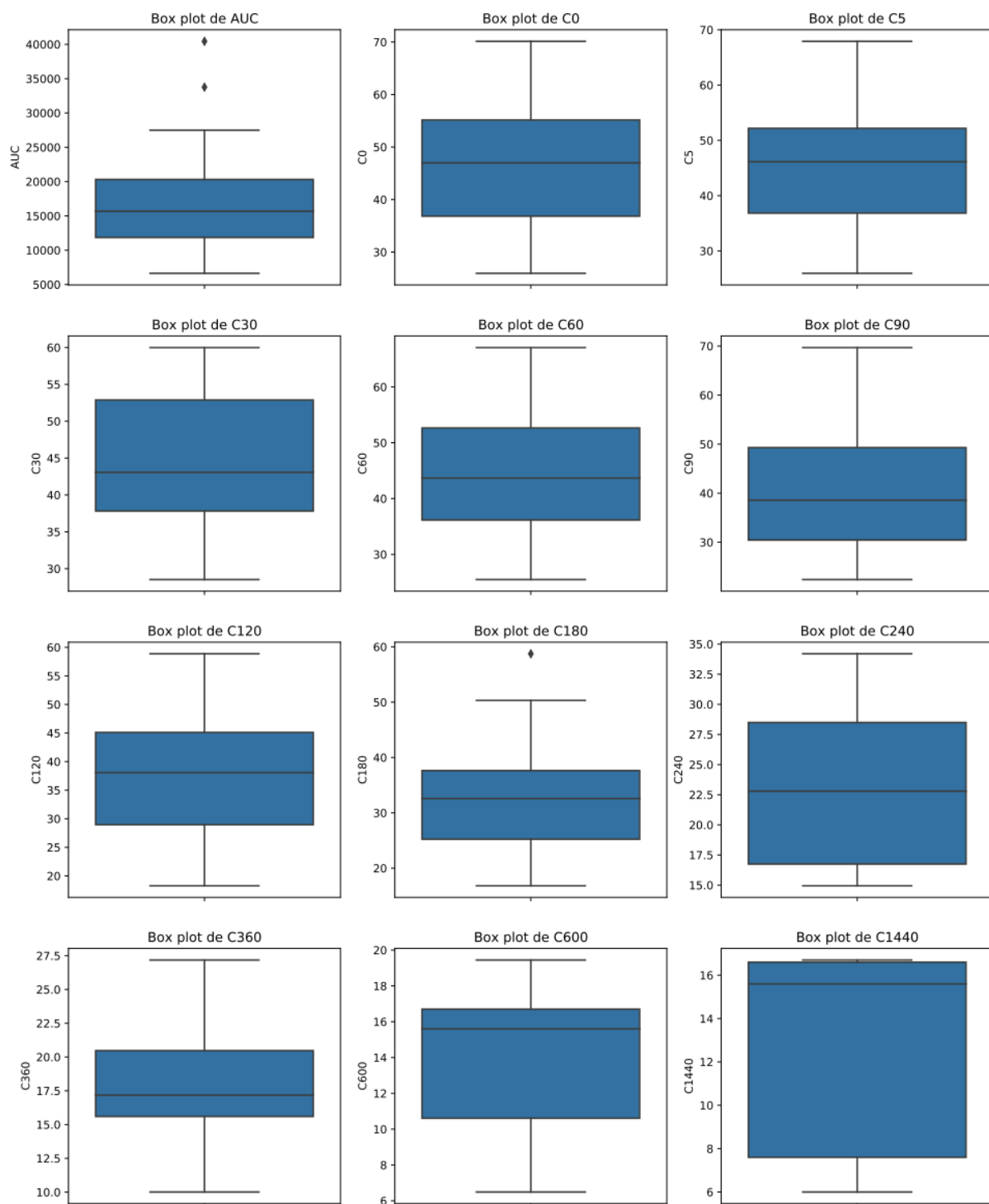
Annexe 14 : Code LSS 3 points

Annexe 15 : Code pour la création des datasets des meilleures composantes principales de la LSS 3 points

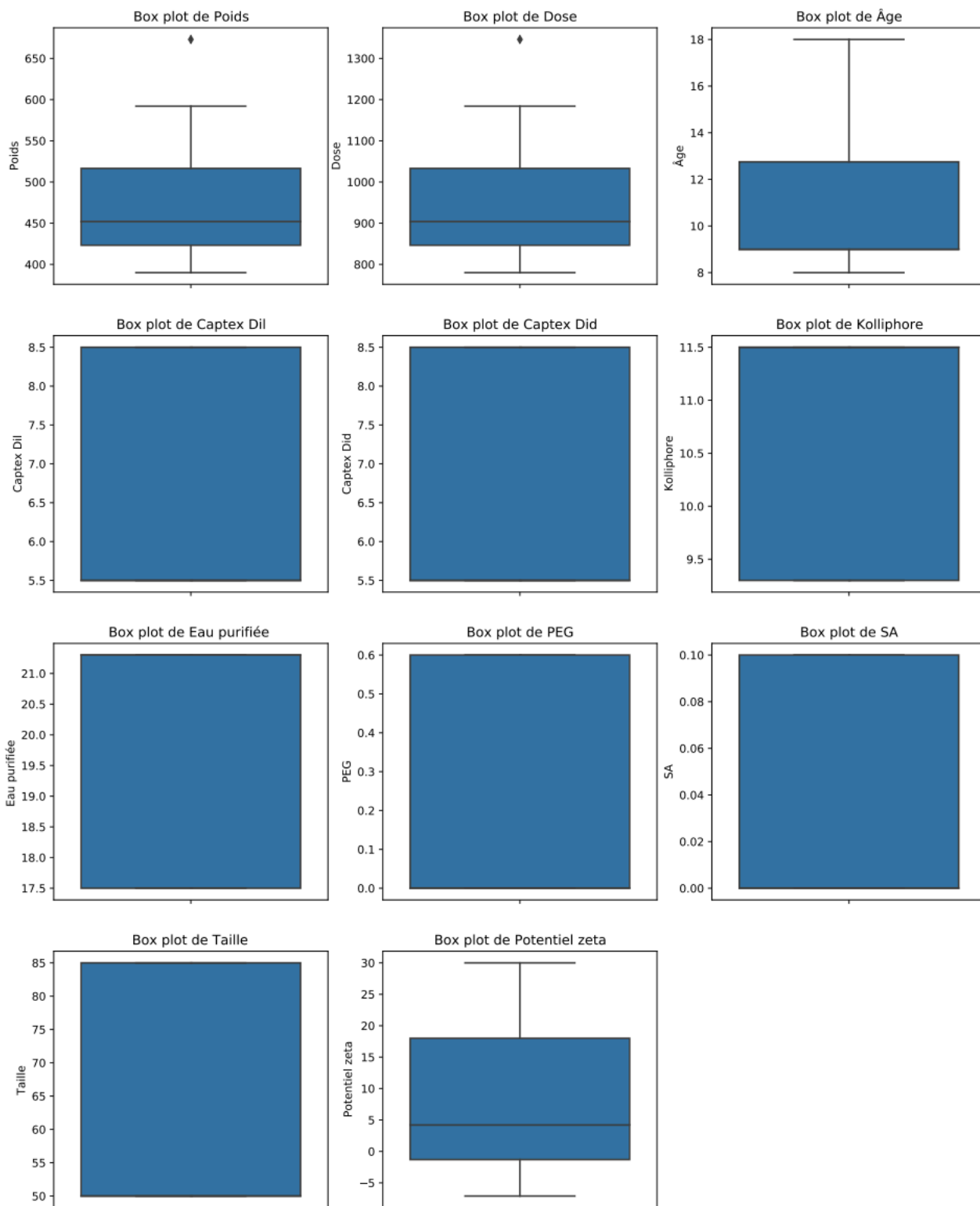
Annexe 16 : Code pour l'amélioration du meilleur modèle de la LSS 3 points

Annexe 17 : Code pour la validation du modèle de LSS 3 points

Annexe 1 (1/2) : Box Plots des variables



Annexe 1 (2/2) : Box Plots des variables



Annexe 2 : Tableau des coefficients des composantes principales ainsi que leurs intervalles de confiance à 95%.

Variables	Valeurs moyennes	P-value maximale	Intervalle de confiance conservateur à 95%
Intercept	16905,22	$4,19 \times 10^{-26}$	[16377,92 ; 17422,47]
PC1	1881,477	$2,59 \times 10^{-14}$	[1652,096 ; 2143,937]
PC3	991,2791	$4,19 \times 10^{-6}$	[622,7701 ; 1364,276]
PC4	475,9979	0,02	[65,33884 ; 871,9915]
PC5	789,3166	0,0009	[333,8557 ; 1242,758]
PC6	-2509,42	$2,99 \times 10^{-8}$	[-3187,63 ; -1869,38]
PC9	-1492,27	0,005	[-2804,83 ; -568,169]
PC10	-7583,06	$7,62 \times 10^{-9}$	[-9455,97 ; -5863,33]
PC11	2689,29	0,03	[229,4892 ; 5247,229]
PC12	5588,376	0,003	[1630,325 ; 8836,399]
PC14	11133,48	0,017	[1534,514 ; 17820,83]
PC18	-39619,5	0,004	[-68334,5 ; -11384,3]
PC19	78544,46	0,002	[26265,65 ; 122075,8]

Annexe 3 (1/2) : Composition des composantes principales sélectionnées pour le modèle de la LSS à 3 points

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
C180	0,20106	0,219572	-0,21494	0,067846	-0,16834	0,375659	0,12224	0,207352	-0,01099	-0,12722	0,002545	-0,13073
C360	0,251363	-0,05452	-0,28492	-0,20674	0,077943	-0,37903	-0,04836	-0,05755	0,031254	-0,32256	0,08705	0,151803
C600	0,122789	-0,3004	0,176622	0,115716	-0,16954	-0,15656	0,147994	0,376509	-0,0609	0,03247	0,307661	0,303553
Poids	-0,13849	-0,06719	-0,34088	0,231142	-0,45217	-0,14023	0,067803	-0,13567	0,725387	0,13651	0,075045	-0,06028
Âge	-0,10865	-0,02635	-0,28811	0,295473	-0,53398	-0,16072	-0,1859	-0,19496	-0,6506	0,057111	-0,03693	-0,05367
Captex Dil	0,166569	0,29886	0,186038	0,012723	-0,1111	-0,16285	-0,02697	-0,00655	0,029487	0,024078	0,031985	-0,02382
Captex Did	0,166569	0,29886	0,186038	0,012723	-0,1111	-0,16285	-0,02697	-0,00655	0,029487	0,024078	0,031985	-0,02382
Kolliphore	-0,16657	-0,29886	-0,18604	-0,01272	0,111104	0,162849	0,026971	0,006548	-0,02949	-0,02408	-0,03198	0,023817
Eau purifiée	-0,16657	-0,29886	-0,18604	-0,01272	0,111104	0,162849	0,026971	0,006548	-0,02949	-0,02408	-0,03198	0,023817
PEG	0,15304	-0,13206	0,115657	-0,3462	-0,29286	0,209668	-0,68324	0,141304	0,120393	0,040813	-0,03409	-0,11727
SA	-0,13447	0,107264	-0,19532	0,388527	0,313034	-0,1034	-0,57065	0,442397	0,079547	-0,06875	0,052427	0,017273
Taille	0,166569	0,29886	0,186038	0,012723	-0,1111	-0,16285	-0,02697	-0,00655	0,029487	0,024078	0,031985	-0,02382
Potentiel zeta	-0,10724	0,26733	-0,06355	0,397566	0,249633	-0,1574	0,078636	0,085139	-0,0095	-0,01258	-0,01093	-0,10625
C180_x_C360	0,271839	0,067478	-0,29033	-0,11372	0,082333	0,030134	0,046553	0,157253	-0,06659	0,460222	0,205629	-0,17951
C180_x_C600	0,285634	-0,06006	-0,0065	0,22756	0,076899	0,240524	-0,09235	-0,2694	0,002849	-0,07057	0,422148	0,121614
C360_x_C600	0,24415	-0,20464	0,117038	0,191791	0,029835	-0,12469	-0,10679	-0,09599	0,055963	0,017509	-0,3696	0,354987
log_C180_x_C360	0,253958	0,04203	-0,33895	-0,16788	0,115846	-0,02396	0,123867	0,187714	-0,08	0,431462	-0,13893	0,006502
log_C180_x_C600	0,271132	-0,14727	0,048146	0,205897	0,058109	0,182853	0,022061	-0,11917	-0,00983	-0,30204	0,291835	-0,41883
log_C360_x_C600	0,206815	-0,2586	0,126001	0,164078	-0,01555	-0,15286	0,046671	0,110534	0,04156	-0,01289	-0,45265	-0,30691
log_C180	0,194251	0,21709	-0,19625	0,078363	-0,21339	0,355227	0,158719	0,265613	0,02101	-0,37015	-0,29472	0,344282
log_C360	0,227901	-0,06808	-0,34004	-0,23657	0,101613	-0,36817	-0,03628	-0,09791	-0,01029	-0,34114	0,027447	-0,05207
log_C600	0,105308	-0,30706	0,166465	0,104677	-0,16838	-0,15697	0,193665	0,438514	-0,04314	0,005848	0,1831	-0,08927
C180_x_C360_x_C600	0,286611	-0,06548	-0,00305	0,229034	0,116958	0,106194	-0,12326	-0,26671	0,032543	0,309297	0,041039	0,396479
log_C180_x_C360_x_C600	0,282523	-0,12519	0,024939	0,210565	0,10464	0,061993	0,008486	-0,14364	0,034646	0,0603	-0,30387	-0,32241

Annexe 3 (2/2) : Composition des composantes principales sélectionnées pour le modèle de la LSS à 3 points

	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24
C180	0,486266	0,267036	-0,38004	0,063295	0,10727	0,23904	0,165926	0,217675	-3,1E-15	1,41E-15	7,8E-16	-2,2E-16
C360	0,177313	0,051614	0,409506	0,349704	-0,25367	0,172383	0,226153	0,222679	-1,5E-15	1,38E-15	1,29E-15	4,3E-16
C600	-0,06562	0,133092	-0,31428	0,197891	-0,2566	-0,31823	-0,23007	0,248226	-1,6E-15	-4,3E-16	1,14E-15	1,79E-16
Poids	-0,06882	0,022183	0,003212	-0,00621	0,005041	-0,0121	-0,0029	-0,00319	8,9E-17	-4,6E-17	-5E-17	8,61E-17
Âge	-0,03449	0,003976	0,037418	0,012932	0,002151	-0,00364	0,007686	0,004408	-2,4E-17	4,82E-17	-4,4E-18	5E-17
Captex Dil	-0,03262	0,006696	0,004946	0,000692	0,055713	-0,03042	0,01587	0,00875	0,687315	0,161229	-0,26407	-0,48153
Captex Did	-0,03262	0,006696	0,004946	0,000692	0,055713	-0,03042	0,01587	0,00875	-0,60941	0,563667	-0,32063	-0,08995
Kolliphore	0,03262	-0,0067	-0,00495	-0,00069	-0,05571	0,030423	-0,01587	-0,00875	0,270016	0,807574	0,254414	0,100942
Eau purifiée	0,03262	-0,0067	-0,00495	-0,00069	-0,05571	0,030423	-0,01587	-0,00875	0,08401	-0,02275	-0,8727	0,17554
PEG	-0,0491	-0,12826	-0,09614	-0,02318	-0,32406	0,188628	-0,08582	-0,0415	1,29E-16	-5,2E-17	-4,9E-16	4,67E-17
SA	-0,03407	0,064366	0,055881	0,057132	0,293686	-0,17185	0,081681	0,055043	-4,3E-16	-5,4E-17	5,04E-16	2,7E-16
Taille	-0,03262	0,006696	0,004946	0,000692	0,055713	-0,03042	0,01587	0,00875	0,276125	0,059924	-0,03359	0,84796
Potentiel zeta	0,000157	-0,16751	-0,16639	-0,10707	-0,63449	0,365729	-0,16843	-0,10414	3,97E-16	6,18E-17	-1,1E-15	-2,7E-16
C180_x_C360	0,306158	0,159892	0,319106	-0,08679	-0,1038	-0,24023	-0,30725	-0,32779	2,91E-15	-2,1E-15	-1,5E-15	-1E-17
C180_x_C600	-0,02679	-0,10545	-0,06239	-0,28711	-0,20638	-0,25904	0,535677	-0,15913	1,71E-15	-2,8E-16	-1,1E-16	2,27E-16
C360_x_C600	0,005732	0,592346	-0,04969	-0,12599	0,000783	0,208801	0,001419	-0,36341	2,41E-15	-4,7E-16	-1,8E-15	-7E-16
log_C180_x_C360	-0,58076	0,034763	-0,13667	-0,06374	0,007279	0,158627	0,283533	0,217606	-1,5E-15	1,77E-15	9,16E-16	-2,1E-16
log_C180_x_C600	-0,38573	0,266461	0,200909	-0,0325	0,113878	0,199882	-0,34195	0,160071	-8,3E-16	9,41E-16	3,51E-16	-3,7E-16
log_C360_x_C600	0,215226	-0,11494	0,158717	-0,43842	-0,10512	-0,2399	0,083997	0,394945	-2,6E-15	8,62E-16	1,99E-15	5,29E-16
log_C180	-0,21308	-0,23892	0,258613	-0,0406	-0,05666	-0,16894	-0,1781	-0,1603	2,13E-15	-1,3E-15	-5,3E-16	2,87E-17
log_C360	0,015812	-0,19855	-0,48707	-0,25963	0,213783	-0,09365	-0,24567	-0,18963	9,8E-16	-1,3E-15	-1,1E-15	-3,7E-16
log_C600	0,061469	-0,31982	0,12683	-0,02193	0,256739	0,405059	0,224726	-0,35588	2,06E-15	2,01E-16	-1,9E-15	-3E-16
C180_x_C360_x_C600	0,188507	-0,38996	0,025015	0,026441	0,236583	0,261002	-0,30329	0,298674	-2,7E-15	1,18E-15	9,41E-16	2,88E-17
log_C180_x_C360_x_C600	-0,03388	-0,18732	-0,18512	0,667583	-0,0258	-0,21823	0,066071	-0,23259	1,37E-15	-2E-15	-7,5E-16	4,45E-16

Annexe 4 (1/4) : Composition du jeu de données initial

ID	AUC	C0	C5	C30	C60	C90	C120	C180	C240	C360	C600
IV-9	21342,85	6,30E+01	6,25E+01	5,96E+01	5,44E+01	6,97E+01	4,69E+01	3,55E+01	2,85E+01	2,06E+01	
IV-11	19779,92	37,45	37,4456435	50,7948357	51,0570861	41,9205031	46,1157742	32,9181695	28,6593863	20,0256858	
IV-12	17953,15	33,21	33,2093472	37,8207007	31,4735328	30,7277902	41,7672908	29,8245424	22,8304706	20,8840796	
RIV-1_w	14921,02	32,22	32,217786	50,55916	36,3939388	26,7640168		23,0983817	18,6875909		1,67E+01
RIV-2_w	8543,13	32,16	32,1624958	33,1542429	28,6717716	30,0061686	21,9010205	17,1192419			1,67E+01
RIV-3_w	25492,19	28,83	28,8323402	34,902549	35,4705427	45,1638825	31,0058765	24,7095707	23,6002728	19,7513723	1,67E+01
IV-PEG-10	10098,58	48,23	48,2263473	52,1172063	40,3515701	38,9835392	28,2010929	21,2140565	15,6	15,6	1,56E+01
IV-PEG-11	8906,33	70,13	67,9133166	57,8355375	40,5851755	31,4486472	30,4089748	16,8209547	15,6	15,6	1,56E+01
IV-PEG-12	7627,07	65,91	64,8741506	5,99E+01	4,13E+01	34,3440252	23,0054703		15,6	15,6	1,56E+01
RIV-PEG-1_w	6633	44,19	43,0605556	37,8203454	38,6793227	22,6343841	19,6733144		15,6	15,6	15,6
RIV-PEG-2_w	11781,7	55,61	52,3501236	3,87E+01	33,4043579	29,4997526	22,087524	25,7718065	15,6	15,6	15,6
RIV-PEG-3_w	6655,49	64,15	56,9877385	31,5369395	25,5010922	22,3928411	18,2718492		15,6	15,6	15,6
IV-SA-5	18448,94	41,1	40,9372064	40,1534319	38,2673165		38,5083432	30,9211161	2,86E+01	1,94E+01	
IV-SA-6	15979,97	49,77	49,7669474	5,41E+01	52,432118	5,09E+01	3,87E+01	3,09E+01	2,51E+01	1,51E+01	
IV-SA-7	15517,98	46,76	4,62E+01	4,32E+01		3,75E+01	2,86E+01	2,95E+01	2,18E+01	1,59E+01	
RIV-SA-1_w	10752,17	33,33	3,29E+01	3,10E+01	31,9676277	3,04E+01		2,44E+01	1,58E+01		13,1
RIV-SA-2_w	15563,1	25,95	25,9512732	3,07E+01		3,86E+01		2,43E+01	2,05E+01	1,56E+01	13,1

Annexe 4 (2/4) : Composition du jeu de données initial

ID	AUC	C0	C5	C30	C60	C90	C120	C180	C240	C360	C600
RIV-SA-3_w	9867,08	38,86	3,88E+01	3,83E+01	3,21E+01	2,77E+01		1,99E+01	1,49E+01		13,1
CIV-1	12069,47	36,77	3,68E+01	4,20E+01	3,20E+01	3,61E+01	3,77E+01	3,73E+01	1,88E+01	1,12E+01	
CIV-2	16758,69	37,05	3,70E+01	3,88E+01	4,53E+01	2,74E+01	2,21E+01	4,25E+01	3,42E+01	1,89E+01	8,45E+00
CIV-3	15794,57	32,31	3,23E+01	4,29E+01	4,37E+01	3,50E+01	3,12E+01	3,29E+01	1,67E+01	1,85E+01	8,08E+00
CIV-1_w	12589,68	49,52	4,89E+01	4,61E+01	4,36E+01	4,97E+01	4,77E+01	3,05E+01	1,80E+01	1,00E+01	
CIV-3_w	21393,98	47,19	4,70E+01	4,59E+01	5,47E+01	4,93E+01	4,85E+01	4,79E+01	2,70E+01	1,96E+01	1,06E+01
CIV-6_w	16751,19	44,53	4,45E+01	5,23E+01	5,42E+01	3,93E+01	3,00E+01	3,47E+01	2,99E+01	1,56E+01	
CIV-PEG-1	40442,94	50,65	50,6480641	5,31E+01	5,33E+01	4,07E+01	4,54E+01	3,76E+01	2,66E+01	2,49E+01	1,94E+01
CIV-PEG-2	22404,12	47,68	4,77E+01	5,06E+01	5,73E+01	4,65E+01	4,51E+01	4,16E+01	2,55E+01	2,50E+01	
CIV-PEG-3	27497,84	53,8	5,17E+01	4,21E+01	4,94E+01	4,45E+01	3,87E+01	3,55E+01	3,42E+01	2,64E+01	1,71E+01
CIV-PEG-1_w	33762,42	58,17	5,82E+01	6,00E+01	5,92E+01	5,84E+01	4,50E+01	5,03E+01	3,31E+01	2,72E+01	1,90E+01
CIV-PEG-2_w	25637,5	59,83	5,95E+01	5,76E+01	5,67E+01	6,12E+01	4,64E+01	3,77E+01	3,10E+01	2,59E+01	
CIV-PEG-3_w	20480,28	59,66	5,96E+01	5,95E+01	5,08E+01	5,58E+01	4,44E+01	4,23E+01	2,91E+01	2,15E+01	
CIV-SA-1	14448,49	67,51	6,53E+01	5,55E+01	6,70E+01	5,70E+01	4,13E+01	3,26E+01	1,94E+01	1,15E+01	
CIV-SA-2	19794,62	50,74	4,61E+01	2,85E+01	4,88E+01	5,09E+01	5,89E+01	5,87E+01	2,81E+01	1,98E+01	8,08E+00
CIV-SA-3	14865,28	29,67	2,97E+01	3,68E+01	4,51E+01	3,31E+01	3,77E+01	2,76E+01	2,28E+01	1,54E+01	7,03E+00
CIV-SA-1_w	14283,22	40,22	3,96E+01	3,67E+01	3,82E+01	3,02E+01	3,23E+01	4,01E+01	1,86E+01	1,19E+01	6,49E+00
RIV-SA-3_w	9867,08	38,86	3,88E+01	3,83E+01	3,21E+01	2,77E+01		1,99E+01	1,49E+01		13,1

Annexe 4 (3/4) : Composition du jeu de données initial

ID	C1440	Poids	Dose	Âge	Captex Dil	Captex Did	Kolliphore	Eau purifiée	PEG	SA	Taille	Potentiel zeta
IV-9	1,67E+01	5,92E+02	1,18E+03	16	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	0,00E+00	5,00E+01	4,20E+00
IV-11	1,67E+01	550	1,10E+03	18	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	0,00E+00	5,00E+01	4,20E+00
IV-12	1,67E+01	673	1,35E+03	9	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	0,00E+00	5,00E+01	4,20E+00
RIV-1_w	1,67E+01	390	7,80E+02	8	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	0,00E+00	5,00E+01	4,20E+00
RIV-2_w	1,67E+01	395	7,90E+02	8	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	0,00E+00	5,00E+01	4,20E+00
RIV-3_w	1,67E+01	400	8,00E+02	8	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	0,00E+00	5,00E+01	4,20E+00
IV-PEG-10	15,6	500	1,00E+03	14	5,5	5,50E+00	1,15E+01	2,13E+01	6,00E-01	0,00E+00	5,00E+01	- 7,10E+00
IV-PEG-11	15,6	552	1,10E+03	14	5,5	5,50E+00	1,15E+01	2,13E+01	6,00E-01	0,00E+00	5,00E+01	- 7,10E+00
IV-PEG-12	15,6	557	1,11E+03	14	5,5	5,50E+00	1,15E+01	2,13E+01	6,00E-01	0,00E+00	5,00E+01	- 7,10E+00
RIV-PEG-1_w	15,6	445	8,90E+02	9	5,5	5,50E+00	1,15E+01	2,13E+01	6,00E-01	0,00E+00	5,00E+01	- 7,10E+00
RIV-PEG-2_w	15,6	459	9,18E+02	9	5,5	5,50E+00	1,15E+01	2,13E+01	6,00E-01	0,00E+00	5,00E+01	- 7,10E+00
RIV-PEG-3_w	15,6	434	8,68E+02	9	5,5	5,50E+00	1,15E+01	2,13E+01	6,00E-01	0,00E+00	5,00E+01	- 7,10E+00
IV-SA-5	13,1	524	1,05E+03	13	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	1,00E-01	5,00E+01	1,80E+01
IV-SA-6	13,1	522	1,04E+03	15	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	1,00E-01	5,00E+01	1,80E+01
IV-SA-7	13,1	585	1,17E+03	15	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	1,00E-01	5,00E+01	1,80E+01
RIV-SA-1_w	13,1	478	9,56E+02	9	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	1,00E-01	5,00E+01	1,80E+01
RIV-SA-2_w	13,1	424	8,48E+02	9	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	1,00E-01	5,00E+01	1,80E+01

Annexe 4 (4/4) : Composition du jeu de données initial

ID	C1440	Poids	Dose	Âge	Captex Dil	Captex Did	Kolliphore	Eau purifiée	PEG	SA	Taille	Potentiel zeta
RIV-SA-3_w	13,1	465	9,30E+02	9	5,5	5,50E+00	1,15E+01	2,13E+01	0,00E+00	1,00E-01	5,00E+01	1,80E+01
CIV-1	7,6	440	8,80E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	0,00E+00	8,50E+01	1,69E+01
CIV-2	7,6	425	8,50E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	0,00E+00	8,50E+01	1,69E+01
CIV-3	7,6	423	8,46E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	0,00E+00	8,50E+01	1,69E+01
CIV-1_w	7,6	440	8,80E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	0,00E+00	8,50E+01	1,69E+01
CIV-3_w	7,6	423	8,46E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	0,00E+00	8,50E+01	1,69E+01
CIV-6_w	7,6	522	1,04E+03	15	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	0,00E+00	8,50E+01	1,69E+01
CIV-PEG-1	16,6	393	7,86E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	6,00E-01	0,00E+00	8,50E+01	- 1,30E+00
CIV-PEG-2	16,6	415	8,30E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	6,00E-01	0,00E+00	8,50E+01	- 1,30E+00
CIV-PEG-3	16,6	433	8,66E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	6,00E-01	0,00E+00	8,50E+01	- 1,30E+00
CIV-PEG-1_w	16,6	393	7,86E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	6,00E-01	0,00E+00	8,50E+01	- 1,30E+00
CIV-PEG-2_w	16,6	415	8,30E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	6,00E-01	0,00E+00	8,50E+01	- 1,30E+00
CIV-PEG-3_w	16,6	433	8,66E+02	9	8,5	8,50E+00	9,30E+00	1,75E+01	6,00E-01	0,00E+00	8,50E+01	- 1,30E+00
CIV-SA-1	6	488	9,76E+02	12	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	1,00E-01	8,50E+01	3,00E+01
CIV-SA-2	6	495	9,90E+02	12	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	1,00E-01	8,50E+01	3,00E+01
CIV-SA-3	6	465	9,30E+02	12	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	1,00E-01	8,50E+01	3,00E+01
CIV-SA-1_w	6	488	9,76E+02	12	8,5	8,50E+00	9,30E+00	1,75E+01	0,00E+00	1,00E-01	8,50E+01	3,00E+01

Annexe 5 : Code pour la description statistique des données

```
import pandas as pd
df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")
df = df.drop(columns=['ID'])
describe_df = df.describe().transpose()
describe_output_file = r"C:\Users\emac\Desktop\LSS\Statistiques descriptives.xlsx"
describe_df.to_excel(describe_output_file)
```

Annexe 6 : Code pour la création des box plots

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib.backends.backend_pdf import PdfPages

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")
df = df.drop(columns=['ID'])

pdf_output_file = r"C:\Users\emac\Desktop\LSS\Box plots.pdf"
with PdfPages(pdf_output_file) as pdf:
    num_columns = 3 # Nombre de colonnes de boxplots par page
    num_rows = (len(df.columns) + num_columns - 1) // num_columns
    fig, axes = plt.subplots(num_rows, num_columns, figsize=(15, 5 * num_rows))
    axes = axes.flatten()
    for ax, column in zip(axes, df.columns):
        sns.boxplot(y=df[column], ax=ax)
        ax.set_title(f'Box plot de {column}')
    for ax in axes[len(df.columns):]:
        fig.delaxes(ax)
    pdf.savefig(fig)
    plt.close()
```

Annexe 7 : Code pour le calcul des coefficients de corrélations

```
import pandas as pd
from scipy.stats import pearsonr, spearmanr

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")

if 'ID' in df.columns:
    df = df.drop(columns=['ID'])

variables_explicatives = df.columns.tolist()
variables_explicatives.remove('AUC')

def calculate_correlations(df, x, y):
    mask = (~df[x].isnull()) & (~df[y].isnull())
    df_filtered = df[mask]
    pearson_corr, _ = pearsonr(df_filtered[x], df_filtered[y])
    spearman_corr, _ = spearmanr(df_filtered[x], df_filtered[y])
    return pearson_corr, spearman_corr

coefficients_pearson = {}
coefficients_spearman = {}

for var in variables_explicatives:
    pearson_corr, spearman_corr = calculate_correlations(df, var, 'AUC')
    coefficients_pearson[var] = pearson_corr
    coefficients_spearman[var] = spearman_corr

for var in variables_explicatives:
    print(f"Variable explicative : {var}")
    print(f"Coefficient de Pearson : {coefficients_pearson[var]}")
    print(f"Coefficient de Spearman : {coefficients_spearman[var]}")
```

```
print("-----")
```

Annexe 8 : Code LSS 1 point

```
import pandas as pd
import numpy as np
from itertools import combinations
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import LeaveOneOut
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import make_scorer, r2_score
from sklearn.feature_selection import RFECV
from sklearn.impute import SimpleImputer
from sklearn.decomposition import PCA

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")

concentrations = ["C0", "C5", "C30", "C60", "C90", "C120", "C180", 'C240', 'C360',
                  'C600', 'C1440']

covariables = ["Poids", "Dose", "Âge", "Captex Dil", "Captex Did", "Kolliphore", "Eau
purifiée ", "PEG", "SA", "Taille ", "Potentiel zeta"]

combinations_of_concentrations = list(combinations(concentrations, 1))

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
    return mpe_percent
```

```

def calculate_adjusted_r2(r2, n, p):
    return 1 - ((1 - r2) * (n - 1) / (n - p - 1))

def calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent):
    return 2 - r2 - adjusted_r2 + rmse_percent / 100 + abs(mpe_percent) / 100

def rmse_scorer(y_true, y_pred):
    return -np.sqrt(np.mean((y_true - y_pred) ** 2))

scorer = make_scorer(rmse_scorer)

models_metrics = []

for combo in combinations_of_concentrations:
    selected_columns = list(combo) + covariables + ["AUC", "ID"]
    df_subset = df[selected_columns].copy()

    log_col1_name = f"log_{combo[0]}"
    df_subset[log_col1_name] = np.log(df_subset[combo[0]])

    y = df_subset['AUC']
    X = df_subset.drop(['AUC', 'ID'], axis=1)

    imputer = SimpleImputer(strategy='median')
    X_imputed = imputer.fit_transform(X)

    scaler = StandardScaler()
    X_scaled = scaler.fit_transform(X_imputed)

    pca = PCA(n_components=X_scaled.shape[1])
    X_pca = pca.fit_transform(X_scaled)

```

```

linear_model = LinearRegression()

rfecv = RFECV(estimator=linear_model, step=1, cv=LeaveOneOut(),
scoring=scorer)
rfecv.fit(X_pca, y)

X_selected = rfecv.transform(X_pca)
selected_features_pca = [f"PC{i+1}" for i in range(X_selected.shape[1])]

selected_features_all = list(X.columns)

y_pred_list = []

loo = LeaveOneOut()
for train_index, test_index in loo.split(X_selected):
    X_train, X_test = X_selected[train_index], X_selected[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    linear_model.fit(X_train, y_train)
    y_pred = linear_model.predict(X_test)

    y_pred_list.append(y_pred[0])
y_pred_list = np.array(y_pred_list)

r2 = r2_score(y, y_pred_list)
rmse_percent = calculate_rmse_percent(y, y_pred_list)
mpe_percent = calculate_mpe_percent(y, y_pred_list)
adjusted_r2 = calculate_adjusted_r2(r2, len(y), X_selected.shape[1])

metric_J = calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent)

```

```
model_metrics = {  
    'Variables utilisées': selected_features_all,  
    'R2': r2,  
    'R2 ajusté': adjusted_r2,  
    'RMSE (%)': rmse_percent,  
    'MPE (%)': mpe_percent,  
    'Métrique J': metric_J  
}  
models_metrics.append(model_metrics)
```

```
metrics_df = pd.DataFrame(models_metrics)  
output_file = r"C:\Users\emac\Desktop\LSS\Résultats LSS 1 point.xlsx"  
metrics_df.to_excel(output_file, index=False)
```

Annexe 9 : Code pour la création des datasets des meilleures composantes principales de la LSS 1 point

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.decomposition import PCA

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")

variables_interet = ['AUC', 'C360', 'Poids', 'Dose', 'Âge', 'Captex Dil', 'Captex Did',
                    'Kolliphore', 'Eau purifiée ', 'PEG', 'SA', 'Taille ', 'Potential zeta']

df['log_C360'] = np.log(df['C360'])

variables_interet += ['log_C360']

df_subset = df[variables_interet]

output_file_before_pca = r"C:\Users\emac\Desktop\LSS\Data set meilleures variables
LSS 1 point.xlsx"
df_subset.to_excel(output_file_before_pca, index=False)

y = df_subset['AUC']
X = df_subset.drop(['AUC'], axis=1)

imputer = SimpleImputer(strategy='median')
X_imputed = imputer.fit_transform(X)

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_imputed)
```

```
pca = PCA()
X_pca = pca.fit_transform(X_scaled)

pca_columns = [f'PC{i+1}' for i in range(X_pca.shape[1])]
df_pca = pd.DataFrame(X_pca, columns=pca_columns)
df_pca['AUC'] = y.values

output_file_pca = r"C:\Users\emac\Desktop\LSS\Composantes principales meilleures
variables LSS 1 point.xlsx"
df_pca.to_excel(output_file_pca, index=False)
```

Annexe 10 : Code pour l'amélioration du meilleur modèle de la LSS 1 point

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import LeaveOneOut
from sklearn.metrics import r2_score
import statsmodels.api as sm

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\Composantes principales meilleures variables LSS 1 point.xlsx")

X = df.drop(columns=['AUC'])
y = df['AUC']

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
    return mpe_percent

def calculate_adjusted_r2(r2, n, p):
    return 1 - ((1 - r2) * (n - 1) / (n - p - 1))

def calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent):
    return 2 - r2 - adjusted_r2 + rmse_percent / 100 + abs(mpe_percent) / 100
```

```
linear_model = LinearRegression()
```

```
models_metrics = []
```

```
while True:
```

```
    loo = LeaveOneOut()
```

```
    p_values_dict = {col: [] for col in X.columns}
```

```
    coefficients_list = []
```

```
    for train_index, test_index in loo.split(X):
```

```
        X_train, X_test = X.iloc[train_index], X.iloc[test_index]
```

```
        y_train, y_test = y.iloc[train_index], y.iloc[test_index]
```

```
        X_train_with_const = sm.add_constant(X_train)
```

```
        model = sm.OLS(y_train, X_train_with_const).fit()
```

```
        for col in X.columns:
```

```
            p_values_dict[col].append(model.pvalues[col])
```

```
            coefficients_list.append(model.params[1:])
```

```
    max_p_values = {col: max(p_values_dict[col]) for col in X.columns}
```

```
    worst_feature = max(max_p_values, key=max_p_values.get)
```

```
    max_p_value = max_p_values[worst_feature]
```

```
    if max_p_value < 0.05:
```

```
        break
```

```
    X = X.drop(columns=[worst_feature])
```

```
y_pred_list = []
```

```
for train_index, test_index in loo.split(X):
```

```

X_train, X_test = X.iloc[train_index], X.iloc[test_index]
y_train, y_test = y.iloc[train_index], y.iloc[test_index]

linear_model.fit(X_train, y_train)
y_pred = linear_model.predict(X_test)

y_pred_list.append(y_pred[0])

y_pred_list = np.array(y_pred_list)

r2 = r2_score(y, y_pred_list)
rmse_percent = calculate_rmse_percent(y, y_pred_list)
mpe_percent = calculate_mpe_percent(y, y_pred_list)
adjusted_r2 = calculate_adjusted_r2(r2, len(y), X.shape[1])

metric_J = calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent)

model_metrics = {
    'Variables utilisées': X.columns.tolist(),
    'R²': r2,
    'R² ajusté': adjusted_r2,
    'RMSE (%)': rmse_percent,
    'MPE (%)': mpe_percent,
    'Métrique J': metric_J
}
models_metrics.append(model_metrics)
metrics_df = pd.DataFrame(models_metrics)

coefficients_df = pd.DataFrame(coefficients_list)
p_values_df = pd.DataFrame.from_dict(p_values_dict)

X_with_const = sm.add_constant(X)

```

```

global_model = sm.OLS(y, X_with_const).fit()
global_significance_df = pd.DataFrame({
    'R-squared': [global_model.rsquared],
    'Adj. R-squared': [global_model.rsquared_adj],
    'F-statistic': [global_model.fvalue],
    'Prob (F-statistic)': [global_model.f_pvalue]
})

output_file = r"C:\Users\emac\Desktop\LSS\Résultats améliorés LSS 1 point.xlsx"
with pd.ExcelWriter(output_file) as writer:
    coefficients_df.to_excel(writer, sheet_name='Coefficients', index=True)
    p_values_df.to_excel(writer, sheet_name='P-Values', index=True)
    global_significance_df.to_excel(writer, sheet_name='Significativité du modèle',
    index=False)
    metrics_df.to_excel(writer, sheet_name='Métriques', index=False)

```

Annexe 11 : Code LSS 2 points

```
import pandas as pd
import numpy as np
from itertools import combinations
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import LeaveOneOut
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import make_scorer, r2_score
from sklearn.feature_selection import RFECV
from sklearn.impute import SimpleImputer
from sklearn.decomposition import PCA

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")

concentrations = ["C0", "C5", "C30", "C60", "C90", "C120", "C180", 'C240', 'C360',
                  'C600', 'C1440']

covariables = ["Poids", "Dose", "Âge", "Captex Dil", "Captex Did", "Kolliphore", "Eau
purifiée ", "PEG", "SA", "Taille ", "Potentiel zeta"]

combinations_of_concentrations = list(combinations(concentrations, 2))

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
    return mpe_percent

def calculate_adjusted_r2(r2, n, p):
    return 1 - ((1 - r2) * (n - 1) / (n - p - 1))
```

```

def calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent):
    return 2 - r2 - adjusted_r2 + rmse_percent / 100 + abs(mpe_percent) / 100

def rmse_scorer(y_true, y_pred):
    return -np.sqrt(np.mean((y_true - y_pred) ** 2))

scorer = make_scorer(rmse_scorer)

models_metrics = []

for combo in combinations_of_concentrations:
    selected_columns = list(combo) + covariables + ["AUC", "ID"]
    df_subset = df[selected_columns].copy()

    interaction_column_name = f"{combo[0]}_x_{combo[1]}"
    df_subset[interaction_column_name] = df_subset[combo[0]] *
df_subset[combo[1]]

    log_col1_name = f"log_{combo[0]}"
    log_col2_name = f"log_{combo[1]}"
    log_interaction_name = f"log_interaction_{combo[0]}_{combo[1]}"

    df_subset[log_col1_name] = np.log(df_subset[combo[0]])
    df_subset[log_col2_name] = np.log(df_subset[combo[1]])
    df_subset[log_interaction_name] = np.log(df_subset[combo[0]]) *
np.log(df_subset[combo[1]])
    df_subset.drop(columns=['ID'], inplace=True)

X = df_subset.drop(columns=['AUC'])
y = df_subset['AUC']

```

```

imputer = SimpleImputer(strategy='median')
X_imputed = imputer.fit_transform(X)

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_imputed)

pca = PCA(n_components=X_scaled.shape[1])
X_pca = pca.fit_transform(X_scaled)

linear_model = LinearRegression()

rfecv = RFECV(estimator=linear_model, step=1, cv=LeaveOneOut(),
scoring=scorer)
rfecv.fit(X_pca, y)

X_selected = rfecv.transform(X_pca)
selected_features_pca = [f"PC{i+1}" for i in range(X_selected.shape[1])] # Noms
des composantes principales

selected_features_all = list(X.columns)

y_pred_list = []

loo = LeaveOneOut()
for train_index, test_index in loo.split(X_selected):
    X_train, X_test = X_selected[train_index], X_selected[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    linear_model.fit(X_train, y_train)
    y_pred = linear_model.predict(X_test)

    y_pred_list.append(y_pred[0])

```

```
y_pred_list = np.array(y_pred_list)
```

```
r2 = r2_score(y, y_pred_list)
```

```
rmse_percent = calculate_rmse_percent(y, y_pred_list)
```

```
mpe_percent = calculate_mpe_percent(y, y_pred_list)
```

```
adjusted_r2 = calculate_adjusted_r2(r2, len(y), X_selected.shape[1])
```

```
metric_J = calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent)
```

```
model_metrics = {
```

```
    'Variables utilisées': selected_features_all,
```

```
    'R2': r2,
```

```
    'R2 ajusté': adjusted_r2,
```

```
    'RMSE (%)': rmse_percent,
```

```
    'MPE (%)': mpe_percent,
```

```
    'Métrique J': metric_J
```

```
}
```

```
models_metrics.append(model_metrics)
```

```
metrics_df = pd.DataFrame(models_metrics)
```

```
output_file = r"C:\Users\emac\Desktop\LSS\Résultats LSS 2 points.xlsx"
```

```
metrics_df.to_excel(output_file, index=False)
```

Annexe 12 : Code pour la création des datasets des meilleures composantes principales de la LSS 2 points

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.decomposition import PCA

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
    return mpe_percent

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")

variables_interet = ['AUC', 'C180', 'C360', 'Poids', 'Dose', 'Âge', 'Captex Dil', 'Captex
Did', 'Kolliphore', 'Eau purifiée ', 'PEG', 'SA', 'Taille ', 'Potentiel zeta']

df['C180_x_C360'] = df['C180'] * df['C360']
df['log_C180'] = np.log(df['C180'])
df['log_C360'] = np.log(df['C360'])
df['log_interaction_C180_C360'] = np.log(df['C180']) * np.log(df['C360'])

variables_interet += ['C180_x_C360', 'log_C180', 'log_C360',
'log_interaction_C180_C360']

df_subset = df[variables_interet]
```

```
output_file_before_pca = r"C:\Users\emac\Desktop\LSS\Data set meilleures variables  
LSS 2 points.xlsx"
```

```
df_subset.to_excel(output_file_before_pca, index=False)
```

```
y = df_subset['AUC']
```

```
X = df_subset.drop(['AUC'], axis=1)
```

```
imputer = SimpleImputer(strategy='median')
```

```
X_imputed = imputer.fit_transform(X)
```

```
scaler = StandardScaler()
```

```
X_scaled = scaler.fit_transform(X_imputed)
```

```
pca = PCA()
```

```
X_pca = pca.fit_transform(X_scaled)
```

```
pca_columns = [f'PC{i+1}' for i in range(X_pca.shape[1])]
```

```
df_pca = pd.DataFrame(X_pca, columns=pca_columns)
```

```
df_pca['AUC'] = y.values
```

```
output_file_pca = r"C:\Users\emac\Desktop\LSS\Composantes principales meilleures  
variables LSS 2 points.xlsx"
```

```
df_pca.to_excel(output_file_pca, index=False)
```

Annexe 13 : Code pour l'amélioration du meilleur modèle de la LSS 2 points

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import LeaveOneOut
from sklearn.metrics import r2_score
import statsmodels.api as sm

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\Composantes principales meilleures variables LSS 2 points.xlsx")

X = df.drop(columns=['AUC'])
y = df['AUC']

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
    return mpe_percent

def calculate_adjusted_r2(r2, n, p):
    return 1 - ((1 - r2) * (n - 1) / (n - p - 1))

def calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent):
    return 2 - r2 - adjusted_r2 + rmse_percent / 100 + abs(mpe_percent) / 100

linear_model = LinearRegression()
```

```
models_metrics = []
```

```
while True:
```

```
    loo = LeaveOneOut()
```

```
    p_values_dict = {col: [] for col in X.columns}
```

```
    coefficients_list = []
```

```
    for train_index, test_index in loo.split(X):
```

```
        X_train, X_test = X.iloc[train_index], X.iloc[test_index]
```

```
        y_train, y_test = y.iloc[train_index], y.iloc[test_index]
```

```
        X_train_with_const = sm.add_constant(X_train)
```

```
        model = sm.OLS(y_train, X_train_with_const).fit()
```

```
        for col in X.columns:
```

```
            p_values_dict[col].append(model.pvalues[col])
```

```
            coefficients_list.append(model.params[1:])
```

```
    max_p_values = {col: max(p_values_dict[col]) for col in X.columns}
```

```
    worst_feature = max(max_p_values, key=max_p_values.get)
```

```
    max_p_value = max_p_values[worst_feature]
```

```
    if max_p_value < 0.05:
```

```
        break
```

```
    X = X.drop(columns=[worst_feature])
```

```
y_pred_list = []
```

```
for train_index, test_index in loo.split(X):
```

```
    X_train, X_test = X.iloc[train_index], X.iloc[test_index]
```

```

y_train, y_test = y.iloc[train_index], y.iloc[test_index]

linear_model.fit(X_train, y_train)
y_pred = linear_model.predict(X_test)

y_pred_list.append(y_pred[0])

y_pred_list = np.array(y_pred_list)

r2 = r2_score(y, y_pred_list)
rmse_percent = calculate_rmse_percent(y, y_pred_list)
mpe_percent = calculate_mpe_percent(y, y_pred_list)
adjusted_r2 = calculate_adjusted_r2(r2, len(y), X.shape[1])

metric_J = calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent)

model_metrics = {
    'Variables utilisées': X.columns.tolist(),
    'R²': r2,
    'R² ajusté': adjusted_r2,
    'RMSE (%)': rmse_percent,
    'MPE (%)': mpe_percent,
    'Métrique J': metric_J
}
models_metrics.append(model_metrics)

metrics_df = pd.DataFrame(models_metrics)

coefficients_df = pd.DataFrame(coefficients_list)
p_values_df = pd.DataFrame.from_dict(p_values_dict)

X_with_const = sm.add_constant(X)

```

```
global_model = sm.OLS(y, X_with_const).fit()
```

```
global_significance_df = pd.DataFrame({  
    'R-squared': [global_model.rsquared],  
    'Adj. R-squared': [global_model.rsquared_adj],  
    'F-statistic': [global_model.fvalue],  
    'Prob (F-statistic)': [global_model.f_pvalue]  
})
```

```
output_file = r"C:\Users\emac\Desktop\LSS\Résultats améliorés LSS 2 points.xlsx"
```

```
with pd.ExcelWriter(output_file) as writer:
```

```
    coefficients_df.to_excel(writer, sheet_name='Coefficients', index=True)
```

```
    p_values_df.to_excel(writer, sheet_name='P-Values', index=True)
```

```
    global_significance_df.to_excel(writer, sheet_name='Significativité du modèle',  
index=False)
```

```
    metrics_df.to_excel(writer, sheet_name='Métriques', index=False)
```

Annexe 14 : Code LSS 3 points

```
import pandas as pd
import numpy as np
from itertools import combinations
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import LeaveOneOut
from sklearn.preprocessing import StandardScaler,RobustScaler
from sklearn.metrics import make_scorer, r2_score
from sklearn.feature_selection import RFECV
from sklearn.impute import SimpleImputer
from sklearn.decomposition import PCA

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")

concentrations = ["C0", "C5", "C30", "C60", "C90", "C120", "C180", 'C240', 'C360',
'C600', 'C1440']

covariables = ["Poids", "Dose", "Âge", "Captex Dil", "Captex Did", "Kolliphore", "Eau
purifiée ", "PEG", "SA", "Taille ", "Potentiel zeta"]

combinations_of_concentrations = list(combinations(concentrations, 3))

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
```

```

return mpe_percent

def calculate_adjusted_r2(r2, n, p):
    return 1 - ((1 - r2) * (n - 1) / (n - p - 1))

def calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent):
    return 2 - r2 - adjusted_r2 + rmse_percent / 100 + abs(mpe_percent) / 100

def rmse_scorer(y_true, y_pred):
    return -np.sqrt(np.mean((y_true - y_pred) ** 2))

scorer = make_scorer(rmse_scorer)

models_metrics = []

for combo in combinations_of_concentrations:
    selected_columns = list(combo) + covariables + ["AUC", "ID"]
    df_subset = df[selected_columns].copy()

    for pair in combinations(combo, 2):
        interaction_column_name = f"{pair[0]}_x_{pair[1]}"
        df_subset[interaction_column_name] = df_subset[pair[0]] * df_subset[pair[1]]

    for pair in combinations(combo, 2):
        log_interaction_column_name = f"log_{pair[0]}_x_{pair[1]}"
        df_subset[log_interaction_column_name] = np.log(df_subset[pair[0]]) *
np.log(df_subset[pair[1]])

    for column in combo:
        log_col_name = f"log_{column}"
        df_subset[log_col_name] = np.log(df_subset[column])

```

```

triple_interaction_name = f"{combo[0]}_x_{combo[1]}_x_{combo[2]}"
df_subset[triple_interaction_name] = df_subset[combo[0]] * df_subset[combo[1]] *
df_subset[combo[2]]

log_triple_interaction_name = f"log_{combo[0]}_x_{combo[1]}_x_{combo[2]}"
df_subset[log_triple_interaction_name] = np.log(df_subset[combo[0]]) *
np.log(df_subset[combo[1]]) * np.log(df_subset[combo[2]])

df_subset.drop(columns=['ID'], inplace=True)

X = df_subset.drop(columns=['AUC'])
y = df_subset['AUC']

imputer = SimpleImputer(strategy='median')
X_imputed = imputer.fit_transform(X)

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_imputed)

pca = PCA(n_components=X_scaled.shape[1])
X_pca = pca.fit_transform(X_scaled)

linear_model = LinearRegression()

rfecv = RFECV(estimator=linear_model, step=1, cv=LeaveOneOut(),
scoring=scorer)
rfecv.fit(X_pca, y)

X_selected = rfecv.transform(X_pca)
selected_features_pca = [f"PC{i+1}" for i in range(X_selected.shape[1])] # Noms
des composantes principales

selected_features_all = list(X.columns)

```

```

y_pred_list = []

loo = LeaveOneOut()
for train_index, test_index in loo.split(X_selected):
    X_train, X_test = X_selected[train_index], X_selected[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    linear_model.fit(X_train, y_train)
    y_pred = linear_model.predict(X_test)

    y_pred_list.append(y_pred[0])

y_pred_list = np.array(y_pred_list)

r2 = r2_score(y, y_pred_list)
rmse_percent = calculate_rmse_percent(y, y_pred_list)
mpe_percent = calculate_mpe_percent(y, y_pred_list)
adjusted_r2 = calculate_adjusted_r2(r2, len(y), X_selected.shape[1])

metric_J = calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent)

model_metrics = {
    'Variables': selected_features_all, # Utilisation de selected_features_all ici
    'R²': r2,
    'R² ajusté': adjusted_r2,
    'RMSE (%)': rmse_percent,
    'MPE (%)': mpe_percent,
    'Métrique J': metric_J
}
models_metrics.append(model_metrics)

```

```
metrics_df = pd.DataFrame(models_metrics)
output_file = r"C:\Users\emac\Desktop\LSS\Résultats LSS 3 points.xlsx"
metrics_df.to_excel(output_file, index=False)
```

Annexe 15 : Code pour la création des datasets des meilleures composantes principales de la LSS 3 points

```
import pandas as pd
import numpy as np
from itertools import combinations
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import LeaveOneOut
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score
from sklearn.impute import SimpleImputer
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy import stats

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
    return mpe_percent

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\data initial.xlsx")

variables_interet = ['AUC', 'C180', 'C360', 'C600', 'Poids', 'Dose', 'Âge', 'Captex Dil',
                    'Captex Did', 'Kolliphore', 'Eau purifiée ', 'PEG', 'SA', 'Taille ', 'Potential zeta']

for pair in combinations(['C180', 'C360', 'C600'], 2):
```

```

interaction_column_name = f"{pair[0]}_x_{pair[1]}"
df[interaction_column_name] = df[pair[0]] * df[pair[1]]
log_interaction_column_name = f"log_{pair[0]}_x_{pair[1]}"
df[log_interaction_column_name] = np.log(df[pair[0]]) * np.log(df[pair[1]])

for column in ['C180', 'C360', 'C600']:
    log_col_name = f"log_{column}"
    df[log_col_name] = np.log(df[column])

df['C180_x_C360_x_C600'] = df['C180'] * df['C360'] * df['C600']

df['log_C180_x_C360_x_C600'] = np.log(df['C180']) * np.log(df['C360']) *
np.log(df['C600'])

variables_interet += [
    'C180_x_C360', 'C180_x_C600', 'C360_x_C600',
    'log_C180_x_C360', 'log_C180_x_C600', 'log_C360_x_C600',
    'log_C180', 'log_C360', 'log_C600',
    'C180_x_C360_x_C600', 'log_C180_x_C360_x_C600'
]
df_subset = df[variables_interet]

output_file_before_pca = r"C:\Users\emac\Desktop\LSS\Data set meilleures variables
LSS 3 points.xlsx"
df_subset.to_excel(output_file_before_pca, index=False)

y = df_subset['AUC']
X = df_subset.drop(['AUC'], axis=1)

```

```
imputer = SimpleImputer(strategy='median')
X_imputed = imputer.fit_transform(X)

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_imputed)

pca = PCA()
X_pca = pca.fit_transform(X_scaled)
pca_columns = [f'PC{i+1}' for i in range(X_pca.shape[1])]
df_pca = pd.DataFrame(X_pca, columns=pca_columns)
df_pca['AUC'] = y.values
output_file_pca = r"C:\Users\emac\Desktop\LSS\Composantes principales meilleures
variables LSS 3 points.xlsx"
df_pca.to_excel(output_file_pca, index=False)
```

Annexe 16 : Code pour l'amélioration du meilleur modèle de la LSS 3 points

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import LeaveOneOut
from sklearn.metrics import r2_score
import statsmodels.api as sm
from statsmodels.stats.outliers_influence import variance_inflation_factor

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\Composantes principales meilleures
variables LSS 3 points.xlsx")

X = df.drop(columns=['AUC'])
y = df['AUC']

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
    return mpe_percent

def calculate_adjusted_r2(r2, n, p):
    return 1 - ((1 - r2) * (n - 1) / (n - p - 1))

def calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent):
    return 2 - r2 - adjusted_r2 + rmse_percent / 100 + abs(mpe_percent) / 100
```

```

def calculate_vif(X):
    vif_data = pd.DataFrame()
    vif_data['Variable'] = X.columns
    vif_data['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
    return vif_data

linear_model = LinearRegression()

models_metrics = []
coefficients_list = []
intercepts_list = []
lower_conf_intervals_list = []
upper_conf_intervals_list = []
p_values_model_list = []
vif_list = []

while True:
    loo = LeaveOneOut()
    p_values_dict = {col: [] for col in X.columns}
    p_values_dict['Intercept'] = []
    coefficients_dict = {col: [] for col in X.columns}
    coefficients_dict['Intercept'] = []
    lower_conf_intervals_dict = {col: [] for col in X.columns}
    lower_conf_intervals_dict['Intercept'] = []
    upper_conf_intervals_dict = {col: [] for col in X.columns}
    upper_conf_intervals_dict['Intercept'] = []

    for train_index, test_index in loo.split(X):
        X_train, X_test = X.iloc[train_index], X.iloc[test_index]
        y_train, y_test = y.iloc[train_index], y.iloc[test_index]

        # Ajuster le modèle avec statsmodels pour obtenir les p-values

```

```

X_train_with_const = sm.add_constant(X_train)
model = sm.OLS(y_train, X_train_with_const).fit()

for col in X.columns:
    p_values_dict[col].append(model.pvalues[col])
    coefficients_dict[col].append(model.params[col])
    lower_conf_intervals_dict[col].append(model.conf_int().loc[col][0])
    upper_conf_intervals_dict[col].append(model.conf_int().loc[col][1])

p_values_dict['Intercept'].append(model.pvalues['const'])
coefficients_dict['Intercept'].append(model.params['const'])
lower_conf_intervals_dict['Intercept'].append(model.conf_int().loc['const'][0])
upper_conf_intervals_dict['Intercept'].append(model.conf_int().loc['const'][1])

max_p_values = {col: max(p_values_dict[col]) for col in p_values_dict.keys()}

worst_feature = max(max_p_values, key=max_p_values.get)
max_p_value = max_p_values[worst_feature]

if max_p_value < 0.05:
    coefficients_list.append(coefficients_dict)
    lower_conf_intervals_list.append(lower_conf_intervals_dict)
    upper_conf_intervals_list.append(upper_conf_intervals_dict)
    break

if worst_feature != 'Intercept':
    X = X.drop(columns=[worst_feature])

y_pred_list = []
vif_dict = {col: [] for col in X.columns}

for train_index, test_index in loo.split(X):

```

```
X_train, X_test = X.iloc[train_index], X.iloc[test_index]
```

```
y_train, y_test = y.iloc[train_index], y.iloc[test_index]
```

```
linear_model.fit(X_train, y_train)
```

```
y_pred = linear_model.predict(X_test)
```

```
y_pred_list.append(y_pred[0])
```

```
X_train_with_const = sm.add_constant(X_train)
```

```
model = sm.OLS(y_train, X_train_with_const).fit()
```

```
p_values_model_list.append(model.f_pvalue)
```

```
vif_df = calculate_vif(X_train)
```

```
for _, row in vif_df.iterrows():
```

```
    vif_dict[row['Variable']].append(row['VIF'])
```

```
y_pred_list = np.array(y_pred_list)
```

```
r2 = r2_score(y, y_pred_list)
```

```
rmse_percent = calculate_rmse_percent(y, y_pred_list)
```

```
mpe_percent = calculate_mpe_percent(y, y_pred_list)
```

```
adjusted_r2 = calculate_adjusted_r2(r2, len(y), X.shape[1])
```

```
metric_J = calculate_metric_J(r2, adjusted_r2, rmse_percent, mpe_percent)
```

```
model_metrics = {
```

```
    'Variables utilisées': X.columns.tolist(),
```

```
    'R2': r2,
```

```
    'R2 ajusté': adjusted_r2,
```

```
    'RMSE (%)': rmse_percent,
```

```
    'MPE (%)': mpe_percent,
```

```
    'Métrique J': metric_J
```

```

}
models_metrics.append(model_metrics)

metrics_df = pd.DataFrame(models_metrics)

coefficients_df = pd.DataFrame(coefficients_list[0])
lower_conf_intervals_df = pd.DataFrame(lower_conf_intervals_list[0])
upper_conf_intervals_df = pd.DataFrame(upper_conf_intervals_list[0])
p_values_df = pd.DataFrame.from_dict(p_values_dict)

p_values_model_df = pd.DataFrame({
    'Itération': np.arange(1, len(p_values_model_list) + 1),
    'P-Value globale': p_values_model_list
})

vif_df = pd.DataFrame(vif_dict)
vif_df.index.name = 'Itération'

X_with_const = sm.add_constant(X)
global_model = sm.OLS(y, X_with_const).fit()
global_significance_df = pd.DataFrame({
    'R-squared': [global_model.rsquared],
    'Adj. R-squared': [global_model.rsquared_adj],
    'F-statistic': [global_model.fvalue],
    'Prob (F-statistic)': [global_model.f_pvalue]
})

output_file = r"C:\Users\emac\Desktop\LSS\Résultats améliorés LSS 3 points.xlsx"
with pd.ExcelWriter(output_file) as writer:
    coefficients_df.to_excel(writer, sheet_name='Coefficients', index=True)
    lower_conf_intervals_df.to_excel(writer, sheet_name='IC Inférieurs', index=True)
    upper_conf_intervals_df.to_excel(writer, sheet_name='IC Supérieurs', index=True)

```

```
p_values_df.to_excel(writer, sheet_name='P-Values', index=True)
p_values_model_df.to_excel(writer, sheet_name='P-Values Modèle LOOCV',
index=False)
vif_df.to_excel(writer, sheet_name='VIF par Itération', index=True)
global_significance_df.to_excel(writer, sheet_name='Significativité du modèle',
index=False)
metrics_df.to_excel(writer, sheet_name='Métriques', index=False)
```

Annexe 17 : Code pour la validation du modèle de LSS 3 points

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import LeaveOneOut
from sklearn.metrics import r2_score
import statsmodels.api as sm
from statsmodels.stats.stattools import durbin_watson
from statsmodels.stats.diagnostic import het_breuschpagan
from scipy.stats import shapiro

def rmse_scorer(y_true, y_pred):
    return -np.sqrt(np.mean((y_true - y_pred) ** 2))

def calculate_rmse_percent(y_true, y_pred):
    N = len(y_true)
    rmse_percent = 100 * np.sqrt((1/N) * np.sum(((y_pred - y_true) / y_true) ** 2))
    return rmse_percent

def calculate_mpe_percent(y_true, y_pred):
    N = len(y_true)
    mpe_percent = 100 * (1/N) * np.sum((y_pred - y_true) / y_true)
    return mpe_percent

def calculate_adjusted_r2(r2, n, p):
    return 1 - ((1 - r2) * (n - 1) / (n - p - 1))

df = pd.read_excel(r"C:\Users\emac\Desktop\LSS\Composantes principales meilleures variables LSS 3 points.xlsx")
```

```

y = df['AUC']
X =
df.drop(['AUC','PC2','PC7','PC8','PC13','PC15','PC16','PC17','PC20','PC21','PC22','PC23','
PC24','PC25'], axis=1)

linear_model = LinearRegression()

y_pred_list = []
coefs_list = []
intercepts_list = []
significance_list = []
model_significance_list = []
studentized_residuals_list = []

loo = LeaveOneOut()
for train_index, test_index in loo.split(X):
    X_train, X_test = X.iloc[train_index], X.iloc[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    linear_model.fit(X_train, y_train)

    y_pred = linear_model.predict(X_test)
    y_pred_list.append(y_pred[0])

    coefs_list.append(linear_model.coef_)
    intercepts_list.append(linear_model.intercept_)

    X_train_sm = sm.add_constant(X_train)
    model = sm.OLS(y_train, X_train_sm).fit()
    significance_list.append(model.pvalues)
    model_significance_list.append(model.f_pvalue) # Significativité du modèle global

```

```

influence = model.get_influence()
studentized_residuals = influence.resid_studentized_external
studentized_residuals_list.append(studentized_residuals[-1])

y_pred_list = np.array(y_pred_list)
studentized_residuals_list = np.array(studentized_residuals_list)

r2 = r2_score(y, y_pred_list)
adjusted_r2 = calculate_adjusted_r2(r2, len(y), X.shape[1])
rmse_percent = calculate_rmse_percent(y, y_pred_list)
mpe_percent = calculate_mpe_percent(y, y_pred_list)

mean_diff = np.mean(y - y_pred_list)
std_diff = np.std(y - y_pred_list, ddof=1)
plt.figure(figsize=(10, 6))
plt.scatter((y + y_pred_list) / 2, y - y_pred_list)
plt.axhline(mean_diff, color='gray', linestyle='--')
plt.axhline(mean_diff + 1.96 * std_diff, color='red', linestyle='--')
plt.axhline(mean_diff - 1.96 * std_diff, color='red', linestyle='--')
plt.xlabel('Mean of observed and predicted')
plt.ylabel('Difference between observed and predicted')
plt.title('Bland-Altman Plot')
plt.show()

plt.figure(figsize=(10, 6))
plt.scatter(y, y_pred_list)
plt.plot([min(y), max(y)], [min(y), max(y)], color='red')
plt.xlabel('Observed AUC')
plt.ylabel('Predicted AUC')
plt.title('Predicted vs Observed')
plt.show()

```

```

residuals = y - y_pred_list

dw_stat = durbin_watson(residuals)
print(f'Statistique de Durbin-Watson : {dw_stat}')
if dw_stat < 1.5 or dw_stat > 2.5:
    print('Indique une possible autocorrélation des résidus.')
else:
    print('Pas d\'autocorrélation des résidus détectée.')

plt.figure(figsize=(10, 6))
sns.scatterplot(x=y_pred_list, y=residuals)
plt.axhline(0, color='black', linestyle='--')
plt.xlabel('Predicted AUC')
plt.ylabel('Residuals')
plt.title('Graphique des résidus')
plt.show()

plt.figure(figsize=(10, 6))
sns.scatterplot(x=y_pred_list, y=studentized_residuals_list)
plt.axhline(0, color='black', linestyle='--')
plt.xlabel('Predicted AUC')
plt.ylabel('Studentized Residuals')
plt.title('Graphique des résidus studentisés')
plt.show()

shapiro_stat, shapiro_p_value = shapiro(residuals)
print(f'Test de Shapiro-Wilk : Statistique={shapiro_stat}, p-value={shapiro_p_value}')
if shapiro_p_value < 0.05:
    print('Les résidus ne suivent pas une distribution normale.')
else:
    print('Les résidus suivent une distribution normale.')

```

```

X_with_const = sm.add_constant(X)
bp_test = het_breuschpagan(residuals, X_with_const)
bp_labels = ['Lagrange multiplier statistic', 'p-value', 'f-value', 'f p-value']
bp_results = dict(zip(bp_labels, bp_test))

print(f"Test de Breusch-Pagan :")
print(f"Statistique : {bp_results['Lagrange multiplier statistic']}")
print(f"Valeur p : {bp_results['p-value']}")
print(f"Statistique F : {bp_results['f-value']}")
print(f"Valeur p de F : {bp_results['f p-value']}")

if bp_results['p-value'] < 0.05:
    print("Les résidus montrent une hétéroscédasticité significative (rejeter H0).")
else:
    print("Les résidus ne montrent pas d'hétéroscédasticité significative (ne pas rejeter H0).")

results_df = pd.DataFrame({
    'Coefficients': coefs_list,
    'Intercept': intercepts_list,
    'Significance': significance_list,
    'Model Significance': model_significance_list,
    'Studentized Residuals': studentized_residuals_list
})

metrics_df = pd.DataFrame({
    'R²': [r2],
    'R² ajusté': [adjusted_r2],
    'RMSE (%)': [rmse_percent],
    'MPE (%)': [mpe_percent]
})

```

```
output_file = r"C:\Users\emac\Desktop\LSS\Validation résultats LSS 3 points.xlsx"
with pd.ExcelWriter(output_file) as writer:
    results_df.to_excel(writer, sheet_name='Coefficients et Significance', index=False)
    metrics_df.to_excel(writer, sheet_name='Metrics', index=False)
```

Table des illustrations

- Figure 1 : Comparaison des tailles des nanomédicaments.
- Figure 2 : Représentation de dendrimères encapsulant un principe actif
- Figure 3 : Illustration de la formation d'une micelle
- Figure 4 : Représentation d'un liposome
- Figure 5 : Représentation schématique d'une nanocapsule lipidique (NCL)
- Figure 6 : Représentation d'un ADC
- Figure 7 : Schéma illustrant la formation des nab-drugs
- Figure 8 : Schéma illustrant la pharmacocinétique des nanomédicaments
- Figure 9 : Représentation de l'impact du phénomène d'opsonisation des nanomédicaments
- Figure 10 : Influence de l'effet EPR sur la distribution des nanomédicaments
- Figure 11 : Schéma illustrant le système cardiovasculaire et le système lymphatique
- Figure 12 : Les différentes voies d'élimination des nanomédicaments
- Figure 13 : Schéma illustrant la dégradation et l'élimination des nanomédicaments par le système SPM
- Figure 14 : Représentation de la filtration glomérulaire, de la sécrétion tubulaire et de la réabsorption tubulaire des médicaments
- Figure 15 : Les différents modes d'internalisation des nanomédicaments
- Figure 16 : Impact de la taille sur la pharmacocinétique des nanomédicaments
- Figure 17 : Différentes formes de nanocapsules
- Figure 18 : Exemple de calcul de l'aire d'un trapèze pour calculer l'AUC entre deux points
- Figure 19 : Les différentes voies pouvant être impliquées dans l'élimination des médicaments
- Figure 20 : Schéma illustrant la notion de volume de distribution
- Figure 21 : Schéma illustrant la technique FRET appliquée aux NCL adapté par le Dr Vincent Lebreton à l'aide de l'article du Dr Émilie Roger
- Figure 22 : Évolution des concentrations plasmatiques des différentes NCL en fonction du temps après injection à J0 (A) et à J7 (B)
- Figure 23 : Exemple de LSS à 3 prélèvements
- Figure 24 : Exemple de nuages de points qui représente la variable Biscuits en fonction de la variable Tea
- Figure 25 : Exemple d'un histogramme d'une variable représentant le taux de taxe d'habitation
- Figure 26 : exemple d'une boîte à moustaches de la variable taux d'habitation
- Figure 27 : Exemple d'utilisation de l'approche bayésienne pour la détermination des paramètres pharmacocinétiques d'un individu
- Figure 28 : Représentation graphique et mathématique du modèle à 1 compartiment avec administration extravasculaire
- Figure 29 : Représentation graphique et mathématique du modèle à 2 compartiments avec administration bolus intravasculaire
- Figure 30 : Représentation graphique et mathématique du modèle à 2 compartiments avec administration extravasculaire
- Figure 31 : Exemple de predicted vs observed
- Figure 32 : Exemple de graphique illustrant l'homoscédasticité (à droite) et l'hétéroscédasticité (à gauche).
- Figure 33 : Exemple de Bland-Altman
- Figure 34 : Exemple de *Visual Predictive Check* (VPC)

Figure 35 : Graphique predicted vs observed

Figure 36 : Graphique des résidus studentisés

Figure 37 : Graphique des résidus

Figure 38 : Tracé de Bland-Altman pour la concordance entre les AUCs mesurées et les AUCs prédites par le modèle de LSS à 3 points

Table des tableaux

Tableau 1 : Composition des NCL formulées

Tableau 2 : Potentiel zêta des différents NCL

Tableau 3 : Tableau des LLOQ des différentes NCL

Tableau 4 : Paramètres pharmacocinétiques des NCL

Tableau 5 : Modèle de population décrivant la pharmacocinétique des NCL chez les rats de type Sprague Dawley

Tableau 6 : Exemple d'une base de données d'effectif $n=10$

Tableau 7 : Exemple des résultats d'un sondage avec valeur manquante auprès de 6 personnes concernant le nombre de médicaments qu'ils prennent par jour

Tableau 8 : Exemple des résultats d'un sondage auprès de 6 personnes concernant le nombre de médicaments qu'ils prennent par jour imputée par la valeur 5

Tableau 9 : Exemple des résultats d'un sondage auprès de 6 personnes concernant le nombre de médicaments qu'ils prennent par jour imputée par la méthode de la moyenne

Tableau 10 : Exemple des résultats d'un sondage auprès de 6 personnes concernant le nombre de médicaments qu'ils prennent par jour imputée par la méthode de la médiane

Tableau 11 : Exemple d'un jeu de données comportant les variables poids et âge

Tableau 12 : Jeu de données comportant les variables poids et âges standardisés

Tableau 13 : Jeu de données comportant les variables poids et âge normalisés

Tableau 14 : Tableau illustrant la sélection de variables par la sélection ascendante

Tableau 15 : Tableau illustrant la sélection de variables par la sélection descendante

Tableau 16 : Tableau explicatif de la sélection de variables par la sélection *stepwise*

Tableau 17 : Tableau comparatif de la régression linéaire multiple et de la modélisation bayésienne

Tableau 18 : Tableau explicatif de la validation croisée K-folds

Tableau 19 : Tableau explicatif de la validation croisée *LOOCV*

Tableau 20 : Tableau illustrant le *bootstrap*

Tableau 21 : Tableau illustrant la RFECV-*LOOCV*

Tableau 22 : Tableau de référencement des données de la littérature sur le choix des seuils des métriques d'évaluation des modèles de LSS

Tableau 23 : Tableau descriptif des statistiques des variables explicatives

Tableau 24 : Tableau présentant les paramètres pharmacocinétiques des différentes NCL

Tableau 25 : Tableau des coefficients de Bravais-Pearson et de Spearman des variables explicatives

Tableau 26 : Tableau des résultats de la LSS à 1 point

Tableau 27 : Tableau des résultats de la LSS à 2 points

Tableau 28 : Tableau des résultats de la LSS à 3 points

Tableau 29 : VIF maximal des variables toutes itérations confondues

Serment de Galien

En présence des Maîtres de la Faculté, je fais le serment :

- D'honorer ceux qui m'ont instruit(e) dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle aux principes qui m'ont été enseignés
- D'actualiser mes connaissances
- D'exercer, dans l'intérêt de la santé publique, ma profession avec conscience et de respecter non seulement la législation en vigueur, mais aussi les règles de Déontologie, de l'honneur, de la probité et du désintéressement ;
- De ne jamais oublier ma responsabilité et mes devoirs envers la personne humaine et sa dignité
- De ne dévoiler à personne les secrets qui m'auraient été confiés ou dont j'aurais eu connaissance dans l'exercice de ma profession
- De faire preuve de loyauté et de solidarité envers mes collègues pharmaciens
- De coopérer avec les autres professionnels de santé En aucun cas, je ne consentirai à utiliser mes connaissances et mon état pour corrompre les mœurs et favoriser des actes criminels.

Que les Hommes m'accordent leur estime si je suis fidèle à mes promesses.

Que je sois couvert(e) d'opprobre et méprisé(e) de mes confrères si j'y manque.

Date

Signature de l'étudiant et du Président du jury

RÉSUMÉ

L'objectif de ce travail était de modéliser la pharmacocinétique de nanocapsules lipidiques et plus précisément de trouver les points les plus prédictifs afin de réduire le nombre de prélèvements sanguins lors des études *in vivo*. Ces nanocapsules ont été développées par le laboratoire de Micro et Nanomédecines Translationnelles à Angers (M.I.N.T). Pour réaliser la modélisation, les nanoparticules de différents types et taille ont été injectées en intraveineux à 34 rats puis à la suite de prélèvements sanguins la concentration de ces dernières a été évaluée à l'aide de la méthode FRET combinée avec un système d'analyse de suivi des nanoparticules. La concentration de ces nanoparticules a été mesurée à différents temps : 5, 30, 60, 90, 120, 180, 240, 360, 600 et 1440 minutes. À la suite de cela, les données ont été collectées et la stratégie d'échantillonnage limitée a été évaluée en utilisant une méthode de régression linéaire multiple sur analyse en composantes principales combinée à une Recursive Feature Elimination-Leave One Out Cross Validation . Ceci a permis de trouver un modèle valide ($r^2 > 0.9$, une RMSE < 15% reflet de la précision et une MPE < 15% reflet du biais) en sélectionnant 3 temps (180, 360, 600 minutes) avec un r^2 de 0.952, un $r^2_{\text{ajusté}}$ égal à 0.925 une MPE égale -0.03 %, une RMSE égale 9.81 % ainsi qu'une métrique J égale à 0.222 relativement proche de 0. Le modèle construit met en évidence la complexité de la pharmacocinétique des nanoparticules sachant qu'elle semble dépendre essentiellement de la phase d'élimination. Cette méthode a permis de conserver uniquement 30% des temps de prélèvements tout en ayant une bonne précision et un faible biais ce qui est prometteur dans le cadre de futurs essais qu'ils soient effectués chez l'animal ou chez l'homme.

Mots-clés : Limited sampling strategies, nanoparticules, régression linéaire, recursive feature elimination, FRET, leave one out cross validation.

Development of a limited sampling strategy for lipid nanocapsules (LNCs)

ABSTRACT

The aim of this work was to model the pharmacokinetics of lipid nanocapsules and, more specifically, to identify the most predictive time points in order to reduce the number of blood samples required in *in vivo* studies. These nanocapsules were developed by the Laboratory of Translational Micro and Nanomedicines in Angers (M.I.N.T). To carry out the modeling, nanoparticles of various types and sizes were administered intravenously to 34 rats. Following blood sampling, the concentration of these nanoparticles was assessed using the FRET method combined with a nanoparticle tracking analysis system. Concentrations were measured at the following time points: 5, 30, 60, 90, 120, 180, 240, 360, 600, and 1440 minutes. The collected data were then used to evaluate a limited sampling strategy through a multiple linear regression method on principal component analysis, combined with Recursive Feature Elimination and Leave-One-Out Cross Validation. This approach enabled the development of a valid model ($r^2 > 0.9$, RMSE < 15% indicating precision, and MPE < 15% reflecting bias) by selecting 3 key time points (180, 360, and 600 minutes). The resulting model showed an r^2 of 0.952, an r^2_{adjusted} of 0.925, an MPE of -0.03%, an RMSE of 9.81%, and a J metric of 0.222, which is relatively close to 0. The model highlights the complexity of nanoparticle pharmacokinetics, which appears to rely predominantly on the elimination phase. This method allowed for a reduction to just 30 percent of the original sampling time points, while maintaining high precision and low bias, which is promising for future studies in both animals and humans.

Keywords: Limited sampling strategies, nanoparticles, linear regression, recursive feature elimination, FRET, leave-one-out cross-validation.