

Université de Lille
Année Universitaire 2026/2027
UFR3S-Pharmacie

**THESE
POUR LE DIPLOME D'ETAT
DE DOCTEUR EN PHARMACIE**

**Soutenue publiquement le 15 septembre 2026
Par Mme. Valentine PIETTRE**

Utilisation du big data en recherche clinique :

Comment les données massives peuvent-elles optimiser la conception, la conduite et les résultats des essais cliniques ?

Membres du jury :

Président : Monsieur Mohamed Lemdani, Professeur des Universités en Biomathématiques, Université de Lille

Directeur de thèse : Professeur Pascal DAO PHAN, Directeur de la Recherche clinique France/Benelux chez Bayer HealthCare SAS à Lille, Professeur Associé de la Faculté de Pharmacie (UFR3S, Université de Lille).

Membre(s) extérieur(s) : Monsieur Jean-Baptiste Colin, Responsable médical oncologie, Bayer Lille

 	LISTE GEREE	LG/FAC/001
Département Pharmacie	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
Document transversal		Page 3/13

Université de Lille

Président
Premier Vice-président
Vice-présidente Formation
Vice-président Recherche
Vice-président Ressources Humaine
Directrice Générale des Services

Régis BORDET
Bertrand DÉCAUDIN
Corinne ROBACZEWSKI
Olivier COLOT
Jean-Philippe TRICOIT
Anne-Valérie CHIRIS-FABRE

UFR3S

Doyen
Premier Vice-Doyen, Vice-Doyen RH, SI et Qualité
Vice-Doyenne Recherche
Vice-Doyen Finances et Patrimoine
Vice-Doyen International
Vice-Doyen Coordination pluriprofessionnelle et Formations sanitaires
Vice-Doyenne Formation tout au long de la vie
Vice-Doyen Territoire-Partenariats
Vice-Doyen Santé numérique et Communication
Vice-Doyenne Vie de Campus
Vice-Doyen étudiant

Dominique LACROIX
Hervé HUBERT
Karine FAURE
Emmanuelle LIPKA
Vincent DERAMECOURT
Sébastien D'HARANCY
Caroline LANIER
Thomas MORGENROTH
Vincent SOBANSKI
Anne-Laure BARBOTIN
Victor HELENA

Faculté de Pharmacie

Vice - Doyen
Premier Assesseur et
Assesseur à la Santé et à l'Accompagnement
Assesseur à la Vie de la Faculté et
Assesseur aux Ressources et Personnels
Responsable de l'Administration et du Pilotage
Représentant étudiant
Chargé de mission 1er cycle
Chargée de mission 2eme cycle
Chargé de mission Accompagnement et Formation à la Recherche
Chargé de mission Relations Internationales
Chargée de Mission Qualité
Chargé de mission dossier HCERES

Pascal ODOU

Anne GARAT

Emmanuelle LIPKA
Cyrille PORTA
Honoré GUISE
Philippe GERVOIS
Héloïse HENRY
Nicolas WILLAND
Christophe FURMAN
Marie-Françoise ODOU
Réjane LESTRELIN

 	LISTE GEREE	LG/FAC/001
Département Pharmacie	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
Document transversal		Page 4/13

Professeurs des Universités - Praticiens Hospitaliers (PU-PH)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
Mme	ALLORGE	Delphine	Toxicologie et Santé publique	81
M.	BROUSSEAU	Thierry	Biochimie	82
M.	DÉCAUDIN	Bertrand	Biopharmacie, Pharmacie galénique et hospitalière	81
M.	DINE	Thierry	Pharmacologie, Pharmacocinétique et Pharmacie clinique	81
Mme	DUPONT-PRADO	Annabelle	Hématologie	82
Mme	GOFFARD	Anne	Bactériologie - Virologie	82
M.	GRESSIER	Bernard	Pharmacologie, Pharmacocinétique et Pharmacie clinique	81
M.	ODOU	Pascal	Biopharmacie, Pharmacie galénique et hospitalière	80
Mme	POULAIN	Stéphanie	Hématologie	82
M.	SIMON	Nicolas	Pharmacologie, Pharmacocinétique et Pharmacie clinique	81
M.	STAELS	Bart	Biologie cellulaire	82

Professeurs des Universités (PU)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M.	ALIOUAT	El Moukhtar	Parasitologie - Biologie animale	87
Mme	ALIOUAT	Cécile-Marie	Parasitologie - Biologie animale	87
Mme	AZAROUAL	Nathalie	Biophysique - RMN	85
M.	BERLARBI	Karim	Physiologie	86
M.	BERTIN	Benjamin	Immunologie	87
M.	BLANCHEMAIN	Nicolas	Pharmacotechnie industrielle	85

 	LISTE GEREE	LG/FAC/001
Département Pharmacie Document transversal	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
		Page 5/13

M.	CARNOY	Christophe	Immunologie	87
M.	CAZIN	Jean-Louis	Pharmacologie, Pharmacocinétique et Pharmacie clinique	86
M.	CUNY	Damien	Sciences végétales et fongiques	87
Mme	DELBAERE	Stéphanie	Biophysique - RMN	85
Mme	DEPREZ	Rebecca	Chimie thérapeutique	86
M.	DEPREZ	Benoît	Chimie bio inorganique	85
Mme	DUMONT	Julie	Biologie cellulaire	87
M.	ELATI	Mohamed	Biomathématiques	27
M.	FOLIGNÉ	Benoît	Bactériologie - Virologie	87
Mme	FOULON	Catherine	Chimie analytique	85
M.	GARÇON	Guillaume	Toxicologie et Santé publique	86
M.	GOOSSENS	Jean-François	Chimie analytique	85
M.	HENNEBELLE	Thierry	Pharmacognosie	86
M.	LEBEGUE	Nicolas	Chimie thérapeutique	86
M.	LEMDANI	Mohamed	Biomathématiques	26
Mme	LESTAVEL	Sophie	Biologie cellulaire	87
Mme	LESTRELIN	Réjane	Biologie cellulaire	87
Mme	LIPKA	Emmanuelle	Chimie analytique	85
Mme	MELNYK	Patricia	Chimie physique	85
M.	MILLET	Régis	Institut de Chimie Pharmaceutique Albert Lespagnol	86
M.	MOREAU	Pierre-Arthur	Sciences végétales et fongiques	87
Mme	MUHR-TAILLEUX	Anne	Biochimie	87
Mme	PERROY	Anne-Catherine	Droit et Economie pharmaceutique	86
Mme	RIVIÈRE	Céline	Pharmacognosie	86

  Université de Lille	LISTE GEREE	LG/FAC/001
Département Pharmacie	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
Document transversal		Page 6/13

Mme	ROMOND	Marie-Bénédicte	Bactériologie - Virologie	87
Mme	SAHPAZ	Sevser	Pharmacognosie	86
M.	SERGHARAERT	Éric	Droit et Economie pharmaceutique	86
M.	SIEPMANN	Juergen	Pharmacotechnie industrielle	85
Mme	SIEPMANN	Florence	Pharmacotechnie industrielle	85
M.	WILLAND	Nicolas	Chimie organique	86

Maîtres de Conférences - Praticiens Hospitaliers (MCU-PH)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
Mme	CUVELIER	Élodie	Pharmacologie, Pharmacocinétique et Pharmacie clinique	81
Mme	DANEL	Cécile	Chimie analytique	85
Mme	DEMARET	Julie	Immunologie	82
Mme	GARAT	Anne	Toxicologie et Santé publique	81
Mme	GENAY	Stéphanie	Biopharmacie, Pharmacie galénique et hospitalière	81
Mme	GILLIOT	Sixtine	Biopharmacie, Pharmacie galénique et hospitalière	80
M.	GRZYCH	Guillaume	Biochimie	82
Mme	HENRY	Héloïse	Biopharmacie, Pharmacie galénique et hospitalière	80
M.	LANNOY	Damien	Biopharmacie, Pharmacie galénique et hospitalière	80
Mme	MASSE	Morgane	Biopharmacie, Pharmacie galénique et hospitalière	81
Mme	ODOU	Marie-Françoise	Bactériologie - Virologie	82

 	LISTE GEREE	LG/FAC/001
Département Pharmacie	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
Document transversal		Page 7/13

Maîtres de Conférences des Universités (MCU)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M.	ANTHÉRIEU	Sébastien	Toxicologie et Santé publique	86
M.	BANTUBUNGI-BLUM	Kadiombo	Biologie cellulaire	87
M.	BERTHET	Jérôme	Biophysique - RMN	85
M	BEDART	Corentin	ICPAL	86
M.	BOCHU	Christophe	Biophysique - RMN	85
M.	BORDAGE	Simon	Pharmacognosie	86
M.	BOSC	Damien	Chimie thérapeutique	86
Mme	BOU KARROUM	Nour	Chimie bioinorganique	
M.	BRIAND	Olivier	Biochimie	87
Mme	CARON-HOUDE	Sandrine	Biologie cellulaire	87
Mme	CARRIÉ	Hélène	Pharmacologie, Pharmacocinétique et Pharmacie clinique	86
Mme	CHABÉ	Magali	Parasitologie - Biologie animale	87
Mme	CHARTON	Julie	Chimie organique	86
M.	CHEVALIER	Dany	Toxicologie et Santé publique	86
Mme	DEMANCHE	Christine	Parasitologie - Biologie animale	87
Mme	DEMARQUILLY	Catherine	Biomathématiques	85
M.	DHIFLI	Wajdi	Biomathématiques	27
M.	EL BAKALI	Jamal	Chimie thérapeutique	86
M.	FARCE	Amaury	Institut de Chimie Pharmaceutique Albert Lespagnol	86
M.	FLIPO	Marion	Chimie organique	86

 Université de Lille Département Pharmacie Document transversal	LISTE GEREE	LG/FAC/001
	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
		Page 8/13

M.	FRULEUX	Alexandre	Sciences végétales et fongiques	
M.	FURMAN	Christophe	Institut de Chimie Pharmaceutique Albert Lespagnol	86
M.	GERVOIS	Philippe	Biochimie	87
Mme	GOOSSENS	Laurence	Institut de Chimie Pharmaceutique Albert Lespagnol	86
Mme	GRAVE	Béatrice	Toxicologie et Santé publique	86
M.	HAMONIER	Julien	Biomathématiques	26
Mme	HAMOUDI-BEN YELLES	Chérifa-Mounira	Pharmacotechnie industrielle	85
Mme	HANNOTHIAUX	Marie-Hélène	Toxicologie et Santé publique	86
Mme	HELLEBOID	Audrey	Physiologie	86
M.	HERMANN	Emmanuel	Immunologie	87
M.	KAMBIA KPAKPAGA	Nicolas	Pharmacologie, Pharmacocinétique et Pharmacie clinique	86
M.	KARROUT	Younes	Pharmacotechnie industrielle	85
Mme	LALLOYER	Fanny	Biochimie	87
Mme	LECOEUR	Marie	Chimie analytique	85
Mme	LEHMANN	Hélène	Droit et Economie pharmaceutique	86
Mme	LELEU	Natascha	Institut de Chimie Pharmaceutique Albert Lespagnol	86
M.	LIBERELLE	Maxime	Biophysique - RMN	
Mme	LOINGEVILLE	Florence	Biomathématiques	26
Mme	MARTIN	Françoise	Physiologie	86
M.	MARTIN MENA	Anthony	Biopharmacie, Pharmacie galénique et hospitalière	
M.	MENETREY	Quentin	Bactériologie - Virologie	87
M.	MORGENROTH	Thomas	Droit et Economie pharmaceutique	86
Mme	MUSCHERT	Susanne	Pharmacotechnie industrielle	85
Mme	NIKASINOVIC	Lydia	Toxicologie et Santé publique	86

 Université de Lille Département Pharmacie Document transversal	LISTE GEREE		LG/FAC/001
	Enseignants et Enseignants-chercheurs 2026-2027		Version 2.3 Applicable au 02/12/2024
			Page 9/13

Mme	PINÇON	Claire	Biomathématiques	85
M.	PIVA	Frank	Biochimie	85
Mme	PLATEL	Anne	Toxicologie et Santé publique	86
M.	POURCET	Benoît	Biochimie	87
M.	RAVAUX	Pierre	Biomathématiques / Innovations pédagogiques	85
Mme	RAVEZ	Séverine	Chimie thérapeutique	86
Mme	ROGEL	Anne	Immunologie	
M.	ROSA	Mickaël	Hématologie	87
M.	ROUMY	Vincent	Pharmacognosie	86
Mme	SEBTI	Yasmine	Biochimie	87
Mme	SINGER	Elisabeth	Bactériologie - Virologie	87
Mme	STANDAERT	Annie	Parasitologie - Biologie animale	87
M.	TAGZIRT	Madjid	Hématologie	87
M.	VILLEMAGNE	Baptiste	Chimie organique	86
M.	WELTI	Stéphane	Sciences végétales et fongiques	87
M.	YOUS	Saïd	Chimie thérapeutique	86
M.	ZITOUNI	Djamel	Biomathématiques	85

Professeurs certifiés

Civ.	Nom	Prénom	Service d'enseignement
Mme	FAUQUANT	Soline	Anglais
M.	HUGES	Dominique	Anglais
Mme	KUBIK	Laurence	Anglais
M.	OSTYN	Gaël	Anglais

 	LISTE GEREE	LG/FAC/001
Département Pharmacie	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
Document transversal		Page 10/13

Professeurs Associés

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M.	BAILLY	Christian	ICPAL	86
M.	DAO PHAN	Haï Pascal	Chimie thérapeutique	86
M.	DHANANI	Alban	Droit et Economie pharmaceutique	86

Maîtres de Conférences Associés

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M	AYED	Elya	Pharmacie officinale	
M.	COUSEIN	Etienne	Biopharmacie, Pharmacie galénique et hospitalière	
Mme	CUCCHI	Malgorzata	Biomathématiques	85
Mme	DANICOURT	Frédérique	Pharmacie officinale	
Mme	DUPIRE	Fanny	Pharmacie officinale	
M.	DUFOSSEZ	François	Biomathématiques	85
M.	FRIMAT	Bruno	Pharmacologie, Pharmacocinétique et Pharmacie clinique	85
Mme	GEILER	Isabelle	Pharmacie officinale	
M.	GILLOT	François	Droit et Economie pharmaceutique	86
M.	MITOUMBA	Fabrice	Biopharmacie, Pharmacie galénique et hospitalière	86
M.	PELLETIER	Franck	Droit et Economie pharmaceutique	86
M	POTHIER	Jean-Claude	Pharmacie officinale	
Mme	ROGNON	Carole	Pharmacie officinale	

 	LISTE GEREE	LG/FAC/001
Département Pharmacie	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
Document transversal		Page 11/13

Assistants Hospitalo-Universitaire (AHU)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M.	BOUDRY	Augustin	Biomathématiques	
Mme	DERAMOUDT	Laure	Pharmacologie, Pharmacocinétique et Pharmacie clinique	
M.	GISH	Alexandr	Toxicologie et Santé publique	
Mme	NEGRIER	Laura	Chimie analytique	

Hospitalo-Universitaire (PHU)

	Nom	Prénom	Service d'enseignement	Section CNU
M.	DESVAGES	Maximilien	Hématologie	
Mme	LENSKI	Marie	Toxicologie et Santé publique	

Attachés Temporaires d'Enseignement et de Recherche (ATER)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
Mme	BERNARD	Lucie	Physiologie	
Mme	BARBIER	Emeline	Toxicologie	
Mme	COMPAGNE	Nina	Chimie Organique	
Mme	COULON	Audrey	Pharmacologie, Pharmacocinétique et Pharmacie clinique	
M.	DUFOSSEZ	Robin	Chimie physique	
Mme	FERRY	Lise	Biochimie	

 	LISTE GEREE	LG/FAC/001
Département Pharmacie	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
Document transversal		Page 12/13

M	HASYEOUI	Mohamed	Chimie Organique	
Mme	HENRY	Doriane	Biochimie	
Mme	KOUAGOU	Yolène	Sciences végétales et fongiques	
M	LAURENT	Arthur	Chimie-Physique	
M.	MACKIN MOHAMOUR	Synthia	Biopharmacie, Pharmacie galénique et hospitalière	
Mme	RAAB	Sadia	Physiologie	

Enseignant contractuel

Civ.	Nom	Prénom	Service d'enseignement
Mme	DELOBEAU	Iris	Pharmacie officinale
M	RIVART	Simon	Pharmacie officinale
Mme	SERGEANT	Sophie	Pharmacie officinale
M.	ZANETTI	Sébastien	Biomathématiques

LRU / MAST

Civ.	Nom	Prénom	Service d'enseignement
Mme	FRAPPE	Jade	Pharmacie officinale
M	LATRON-FREMEAU	Pierre-Manuel	Pharmacie officinale
M.	MASCAUT	Daniel	Pharmacologie, Pharmacocinétique et Pharmacie clinique

 	LISTE GEREE	LG/FAC/001
Département Pharmacie	Enseignants et Enseignants-chercheurs 2026-2027	Version 2.3 Applicable au 02/12/2024
Document transversal		Page 13/13

CYCLE DE VIE DU DOCUMENT

Version	Modifié par	Date	Principales modifications
1.0		20/02/2020	Création
2.0		02/01/2022	Mise à jour
2.1		21/06/2022	Mise à jour
2.2		01/02/2024	Mise à jour
2.3		15/11/2024	Mise à jour
2.4		18/02/2025	Mise à jour

UFR3S - Faculté de Pharmacie de Lille

3 Rue du Professeur Laguesse – 59000 Lille
03 20 96 40 40
<https://pharmacie.univ-lille.fr>

L'Université n'entend donner aucune approbation aux opinions émises dans les thèses ; celles-ci sont propres à leurs auteurs.

Remerciements

Un immense merci à Monsieur le Professeur Pascal Dao Phan, d'avoir accepté de m'accompagner pour la rédaction de cette thèse de pharmacie. Et surtout merci de m'avoir donné l'opportunité d'intégrer le service des Opérations cliniques de Bayer, dans lequel je m'épanouis et grandis professionnellement et humainement depuis plus de 4 ans.

Je souhaite également remercier Monsieur le Professeur Mohamed Lemdani et Monsieur Jean-Baptiste Colin, d'avoir accepté de faire partie de mon jury de soutenance de thèse. Merci à Jean-Baptiste d'avoir en partie inspiré le sujet de cette thèse, grâce à un cours donné il y a quelques années à ILIS, qui m'avait particulièrement passionnée.

Merci également à Monsieur le Doyen de la faculté de pharmacie de Lille, ainsi qu'à l'ensemble du personnel de la faculté, qui ont contribué à mon parcours d'étudiante puis de jeune pharmacienne, et qui ont permis l'aboutissement de cette thèse ainsi que l'organisation de ma soutenance.

Je remercie tout particulièrement mes collègues et amis du service des Opérations cliniques de Bayer, avec qui j'ai la chance de collaborer au quotidien, toujours avec bonne humeur et exigence. Merci également aux managers qui ont accompagné mon apprentissage et mon évolution depuis mon arrivée chez Bayer.

Je tenais également à remercier ma sœur Eugénie, qui a foulé avant moi le sol de la faculté de pharmacie, puis celui du laboratoire Bayer. Merci pour nos discussions passionnantes sur les enjeux de l'accès au marché des médicaments, qui ont nourri mes réflexions pour cette thèse.

Merci à mes parents, de m'avoir donné les moyens de faire les études auxquelles j'aspirais, et de m'avoir soutenue dans les différentes étapes de ma vie scolaire et d'étudiante.

Merci Papy et Mamie, de m'avoir partagé votre passion de la pharmacie, et d'avoir éclairé ce choix de carrière grâce à votre expérience de pharmaciens d'officine.

Enfin, merci à Henri, de me soutenir dans tous mes projets, chacun avec leur lot d'ambition et de défis, mais qui sont toujours de grandes sources de satisfaction pour moi.

Table des matières

I.	Introduction.....	20
1)	La recherche clinique	20
2)	Le big data	22
3)	Objectifs de la thèse.....	23
II.	Limites du modèle actuel des essais cliniques	24
1)	Autorités de santé et enjeux	24
a.	Autorisation de mise sur le marché.....	25
b.	Taux de remboursement des médicaments	25
c.	Prix des médicaments.....	27
2)	Le Gold Standard des essais cliniques	29
3)	Evolution des objectifs et de la conduite des essais cliniques	31
a.	Médecine personnalisée.....	332
b.	Enjeux de la médecine personnalisée.....	33
4)	Exemple du développement du Larotrectinib.....	35
a.	Essais cliniques et résultats.....	36
b.	Accès au marché	40
c.	Obstacles et freins à la mise sur le marché.....	41
III.	Essor du Big Data et de l'Intelligence Artificielle.....	44
1)	Les sources de big data	46
a.	Dossiers médicaux électroniques (DME) et dossiers électroniques du patient (DEP).....	47
b.	Bases de données de santé publique.....	48
c.	Essais cliniques précédents et registres.....	48
d.	Données génomiques	49
e.	Données de surveillance et de pharmacovigilance	50
f.	Données des réseaux sociaux et forums de patients	50
g.	Données de capteurs et de dispositifs médicaux	51
h.	Données issues de la télémédecine	52
IV.	Utilisation du Big Data dans les essais cliniques	53
1)	Accélération des découvertes	53
2)	Optimisation de la conception des essais	56
3)	Essais clinique innovants.....	58
a.	Cohortes synthétiques, ou cohortes externes	58
b.	Bras <i>in silico</i>	61
c.	Jumeaux numériques.....	63
V.	Conclusion	65

Index des abréviations

AMM : Autorisation de Mise sur le Marché
ANSM : Agence Nationale de Sécurité des Médicaments et des Produits de Santé
ASMR : Amélioration du Service Médical Rendu
BPC : Bonnes Pratiques Cliniques
CépiDc (Centre d'épidémiologie sur les causes médicales de décès
CEPS : Comité Économique des Produits de Santé
CHMP : Committee for Medicinal Products for Human Use
CNAM : Caisse Nationale de l'Assurance Maladie
CPNPC : Cancer du Poumon Non à Petites Cellules
CSP : Code de la Santé Publique
CNIL : Commission Nationale de l'Informatique et des Libertés
CT : Commission de la Transparence
DEP : Dossier électronique du patient
DME : Dossier Médical Electronique
DoR : Duration of Response (Durée de réponse)
DPO : Data Privacy Officer
ECG : Electrocardiogramme
EMA : European Medicines Agency (Agence Européenne du Médicament)
FDA : Food and Drugs Administration
HAS : Haute Autorité de Santé
HER2 : Human Epidermal growth factor Receptor 2 (Récepteur du facteur de croissance épidermique humain 2)
HTA : autorités d'évaluation des technologies de santé
HRQoL : Health-Related Quality of Life (Qualité de vie rapportée par les patients)
IA : Intelligence Artificielle
ICER : Incremental Cost-Effectiveness Ratio (Ratio coût-efficacité incrémental)
ICH-GCP : International Council for Harmonization on Good Clinical Practice
MIT : Massachusetts Institute of Technology
NGS : Next Generation Sequencing (Séquençage de Nouvelle Génération)
NHS : National Health Service
ORR : Overall Response Rate (Taux de Réponse Objective)
OS : Overall Survival (Survie Globale)
PD : Pharmacodynamie

PDS : Plateforme des Données de Santé
PFS : Progression-Free Survival (Survie sans Progression)
PK : Pharmacocinétique
PMSI : Programme de Médicalisation des Systèmes d'Information
R&D : Recherche et Développement
RCP : Réunion de Concertation Pluridisciplinaire
RECIST : Response Evaluation Criteria in Solid Tumors
RGPD : Règlement Général sur la Protection des Données
RIPH : Recherche Impliquant la Personne Humaine
RNA : Acide Ribonucléique
SMR : Service Médical Rendu
SNIIRAM : Système National d'Information Inter-Régimes de l'Assurance Maladie
SNDS : Système National des Données de Santé
SpO₂ : Taux de Saturation en Oxygène
TOC : Troubles Obsessionnels Compulsifs
TTP : Time to Progression (Temps jusqu'à progression)

I. Introduction

1) La recherche clinique

Le terme « recherche clinique » englobe les études scientifiques réalisées sur la personne humaine, en vue du développement des connaissances médicales. Ces études impliquent le suivi de patients ou de volontaires sains pour mieux comprendre les maladies et, dans certains cas, pour obtenir des données permettant de commercialiser de nouveaux médicaments ou dispositifs médicaux. Nous nous intéresserons ici particulièrement au cas précis du médicament, où l'objectif final de la recherche est la mise sur le marché d'un nouveau principe actif, ou d'un principe actif existant dans une nouvelle indication. La recherche clinique marque dans ce cas le passage d'une découverte scientifique à une innovation validée pour un usage médical, avec l'objectif de mesurer la balance bénéfice/risque de l'administration du médicament, et de définir les conditions optimales de son utilisation¹. La recherche clinique est aujourd'hui indispensable pour permettre l'innovation médicale.

Elle se divise en 4 grandes phases de recherche.

La phase I est celle de la première administration du traitement à l'Homme. Elle porte sur un nombre restreint de patients ou de volontaires sains, et son objectif est avant tout de déterminer le profil de sécurité du médicament, ainsi que sa pharmacocinétique (effets du corps sur le médicament) et sa pharmacodynamie (effets du médicament sur le corps). Environ 70% des produits expérimentés en phase I passent ensuite en phase II.²

La phase II se déroule sur des cohortes de patients plus larges et commence à s'intéresser à l'efficacité du produit. Les essais de phase II peuvent également servir à déterminer la dose optimale de traitement expérimental à utiliser pour la suite du développement. Environ un tiers des médicaments expérimentaux continuent leur développement vers des essais de phase III.

La phase III quant à elle, inclue un grand, voire très grand, nombre de patients. Cette phase se fait généralement en comparaison avec un autre traitement déjà sur le marché, ou un produit sans principe actif, appelé placebo. A la suite d'un essai de phase III, entre 70 et 90% des nouveaux médicaments sont retenus pour une demande d'autorisation de mise sur le marché (AMM).

Finalement, sur 10 médicaments qui entrent en phase de recherche sur l'Homme, on estime qu'un seul sera finalement mis sur le marché.

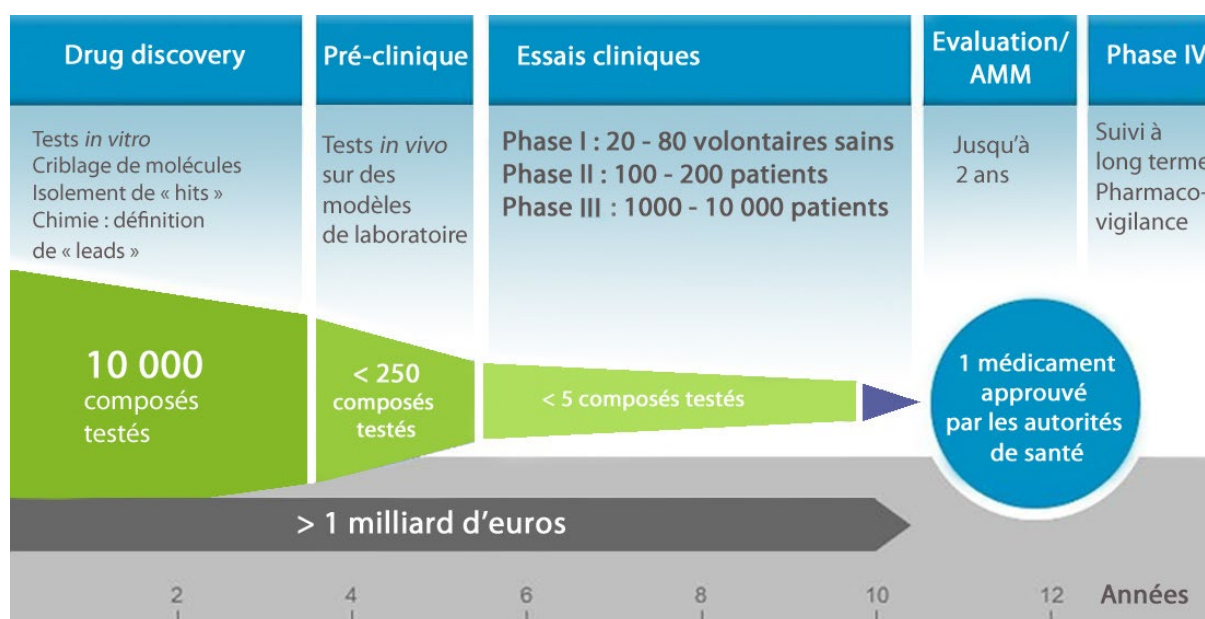


Figure 1. Processus de développement et de découverte de nouveaux médicaments³

Une toute dernière étape de recherche, la phase IV, s'effectue une fois que le traitement est disponible sur le marché, pour une surveillance à long terme et à grande échelle. En effet, lors des études cliniques, les patients sont inclus selon des critères d'inclusion précis, afin d'obtenir une population avec des caractéristiques optimales, permettant d'un côté d'assurer au mieux leur sécurité mais également de potentialiser les probabilités de détecter un effet thérapeutique. Les patients inclus dans un essai clinique présentent généralement des comorbidités limitées et considérées comme peu cliniquement significatives, et la liste des traitements pris de manière concomitante au médicament expérimental est également limitée et contrôlée. C'est donc au cours de la phase IV que des effets indésirables très rares, non reportés en phase I à III, sont détectés. Cette dernière phase garantit le bon usage du médicament et assure la poursuite de la pharmacovigilance du nouveau médicament une fois utilisé en vie réelle.

Bien que très peu encadrée à cette époque, la recherche clinique remonte à l'Antiquité et au Moyen-Age, avec de premières observations des effets de traitements chez des patients. Ce concept s'est ensuite largement précisé, en particulier au cours du siècle dernier. Ainsi, en 2024, ce ne sont pas moins de 834 essais cliniques portant

sur des médicaments qui ont été autorisés par l'Agence Nationale de Sécurité des Médicaments et des Produits de santé (ANSM), et donc menés sur des patients ou volontaires sains français.⁴

Plusieurs textes et organismes régissent aujourd'hui la recherche clinique en France. On peut tout d'abord citer le Règlement Européen n° 536/2014, entré en application en janvier 2022, qui régit les essais « médicaments » menés en Europe, et que les essais menés en France sont donc tenus de suivre. Le Code de la santé publique (CSP), quant à lui, encadre la Recherche impliquant la personne humaine (RIPH) et les obligations nationales françaises. Les Bonnes Pratiques Cliniques, ou ICH-GCP, constituent un standard de conduite pour la recherche clinique. Enfin, la Commission Nationale de l'Informatique et des Libertés (CNIL), impose les règles à suivre pour le traitement des données collectées au cours des essais.

2) Le big data

Le Big Data, quant à lui, est un concept bien plus récent. Ce n'est qu'à partir des années quatre-vingt-dix qu'on commence à entendre parler de Big Data. On définit souvent le Big Data par les « 3 V » qui le caractérisent : le Volume et la Variété des données, et la Vélocité avec laquelle elles sont générées, collectées et traitées.⁵ C'est ce qui les différencie des données traditionnelles.

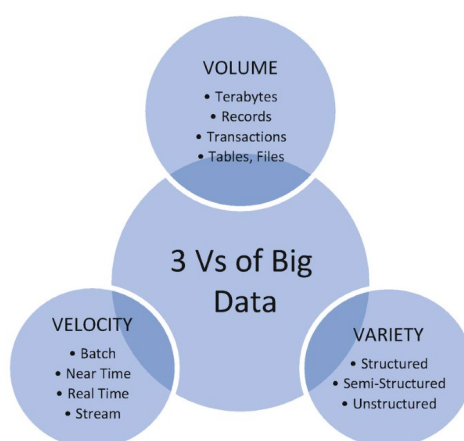


Figure 2. Big Data – Les 3 V⁶

Le monde du Big Data a depuis rapidement évolué. Les tendances émergentes dans ce domaine transforment la manière dont nous analysons et exploitons les

données. L'intelligence artificielle et le machine learning (ou apprentissage automatique, en français) seront probablement au cœur des prochaines révolutions industrielles et scientifiques. Les entreprises doivent alors s'adapter à ces changements pour rester compétitives.⁷ C'est évidemment le cas pour l'industrie pharmaceutique, et en particulier pour le domaine de la recherche clinique.

En effet, l'utilisation du big data dans la recherche clinique est potentiellement en mesure de révolutionner la manière dont les essais sont conçus, exécutés et leurs résultats analysés. Il est en effet possible de récupérer, utiliser et analyser de grandes quantités de données provenant de diverses sources que nous détaillerons par la suite, afin de les intégrer dans les processus de conception et de réalisation des essais cliniques. Les objectifs de l'utilisation du big data en recherche clinique sont multiples : gagner en efficacité, en rapidité, en exhaustivité et en compétitivité, répondre à certaines questions éthiques, et suivre les évolutions actuelles des tendances en matière de recherche clinique.

3) Objectifs de la thèse

Nous verrons au cours de cette présentation que le système actuel d'évaluation des médicaments est confronté à plusieurs défis majeurs, tels que la durée des essais cliniques, les coûts élevés liés au développement des médicaments, les problématiques d'inégalités dans l'accès aux traitements, les enjeux éthiques, ou encore la volonté de développer toujours plus rapidement des médicaments toujours plus ciblés et personnalisés.

Dans ce contexte, le Big Data offre de potentielles solutions pour répondre à ces problématiques, via des approches novatrices, telles que les cohortes synthétiques et les jumeaux numériques. Il est cependant important de rappeler que ces approches novatrices doivent permettre de réduire les contraintes citées précédemment, tout en maintenant des normes élevées de sécurité, d'efficacité et de démonstration des effets.

Ainsi, ces innovations, basées sur la modélisation de populations virtuelles et la simulation de réponses thérapeutiques, pourraient révolutionner l'évaluation des

médicaments en permettant, entre autres, une accélération des délais et une meilleure gestion des coûts de développement.

Cependant, cette transformation soulève plusieurs questions fondamentales auxquelles nous tenterons de répondre au cours de cette thèse. Premièrement, il sera important de comprendre qui sont les acteurs qui doivent accepter ces nouvelles méthodes. Ces acteurs incluent notamment les autorités de santé et les laboratoires pharmaceutiques eux-mêmes. Ensuite, nous évoquerons les conditions nécessaires pour qu'une innovation satisfasse ces acteurs. Il sera alors essentiel d'explorer les preuves scientifiques, les standards réglementaires, et les protocoles technologiques nécessaires pour garantir leur acceptabilité. Enfin, nous parlerons de la capacité de ces acteurs à adapter leurs cadres de décision, leurs outils d'évaluation, et leurs compétences face aux produits et technologies émergents. Nous évoquerons ainsi la capacité d'acceptation actuelle et potentielle des cohortes du Big Data dans le processus de développement des médicaments. Il conviendra de ne pas oublier d'analyser les impacts potentiels sur l'évaluation des médicaments, tant au niveau des délais, des coûts, que de l'accessibilité aux traitements.

II. Limites du modèle actuel des essais cliniques

1) Autorités de santé et enjeux

Les autorités de santé jouent un rôle particulièrement important dans le domaine de la recherche clinique. En effet, ce sont ces autorités qui ont la responsabilité de protéger la santé publique en garantissant l'efficacité et la sécurité des médicaments, des vaccins et autres produits biologiques à usage humain et vétérinaire, et des dispositifs médicaux.

Ces autorités peuvent être nationales, comme la HAS (Haute Autorité de Santé) et l'ANSM, qui sont les autorités responsables en France. Aux Etats-Unis, c'est la Food and Drug Administration (FDA), qui assure l'évaluation des produits de santé.

Il existe également des autorités de santé internationales, comme l'EMA (Agence Européenne du Médicament), qui évaluent les produits de santé de façon globale pour plusieurs pays – ici l'ensemble des pays européens.

a. Autorisation de mise sur le marché

Parmi les missions de ces agences, se trouve l'évaluation des demandes d'AMM des médicaments. En effet, après avoir mené des essais cliniques et collecté des données suffisantes, les entreprises pharmaceutiques réalisent et déposent un dossier de demande de mise sur le marché, censé démontrer la positivité de la balance bénéfice/risque dudit médicament, à partir des données collectées lors des essais cliniques notamment. L'AMM sera délivrée pour un pays donné lorsque les demandes d'autorisation sont faites auprès d'une autorité nationale. L'AMM peut également être délivrées pour plusieurs pays à la fois, lorsque la demande est faite de manière centralisée.

Aujourd'hui, dans la grande majorité des cas, les demandes d'autorisation de mise sur le marché d'un médicament pour les pays européens se fait de manière centralisée, via une demande centralisée auprès de l'EMA. Via son comité scientifique principal, le CHMP (Committee for Medicinal Products for Human Use), l'EMA évalue les données cliniques et précliniques soumises par le laboratoire pour garantir que le médicament est sûr, efficace et de qualité. Une fois un avis favorable émis par le CHMP, la Commission européenne accorde l'AMM, valable dans l'ensemble des pays membres de l'Union européenne.

Les données des essais servent également à la négociation du prix et du taux de remboursement des médicaments. L'objectif pour les entreprises pharmaceutiques est alors, non seulement, d'obtenir une AMM en démontrant la positivité de la balance bénéfice/risque du traitement expérimental, mais également de prouver sa supériorité en comparaison avec le traitement de référence, ou un placebo. Cette négociation se fait une nouvelle fois avec les autorités de santé.

b. Taux de remboursement des médicaments

En France, le système de santé publique, avec la Sécurité sociale, rembourse une partie des frais liés aux médicaments prescrits par les professionnels de santé.

Une commission spécifique de la HAS : La Commission de la Transparence (CT), évalue les résultats des essais cliniques pour décider de l'octroi ou non d'un remboursement pour chaque médicament.

De cette façon, la France garde la main sur les médicaments auxquels les patients, qui dépendent de la sécurité sociale, auront accès. Cela permet de contourner localement les AMM centralisées européennes en posant les conditions de prise en charge particulières pour les patients - par exemple des conditions de sous populations, d'âge, de pathologie spécifique, de gravité, etc. - qui ont la possibilité de préciser l'AMM obtenue de manière centralisée. De cette façon, le gouvernement français est généralement plus restrictif, pour garder la maîtrise des coûts de santé publique français.

La Commission de la Transparence détermine ainsi un score de Service Médical Rendu (SMR)⁸, en fonction de la démonstration d'une efficacité clinique pertinente et d'un profil de tolérance acceptable avec un niveau de preuve suffisant au regard du contexte clinique (population de l'étude, critères de jugement, durée de l'étude ...). Le SMR peut être qualifié en 3 niveaux :

- Important : permettant d'obtenir un remboursement de 65%
- Modéré : permettant d'obtenir un remboursement à 30%
- Faible : permettant d'obtenir un remboursement à 15%

Mais certains facteurs peuvent conduire à un SMR insuffisant, et il n'y aura alors pas de remboursement par la Solidarité Nationale. Cela ne signifie pas pour autant que le médicament est inefficace, ni qu'il ne sera pas mis sur le marché.

Par ailleurs, les médicaments jugés comme irremplaçables et particulièrement coûteux peuvent bénéficier d'une prise en charge à 100% par le système de santé publique.

Catégorie de médicaments	Taux de remboursement
Médicament irremplaçable pour affections graves et invalidantes	100%
Médicament à SMR majeur ou important et préparations magistrales	65%
Médicament à SMR modéré	30%
Médicament à SMR faible	15%
Médicament à SMR insuffisant	Pas de remboursement

Figure 3. Taux de remboursement en fonction du type de médicament⁹

c. Prix des médicaments

La Commission de la Transparence détermine également un score d'amélioration du service rendu (ASMR), en fonction des médicaments déjà sur le marché, considérés comme traitement de référence (ou « standard of care », en anglais) pour la pathologie concernée, s'il en existe.

L'objectif des essais cliniques est donc de déterminer, dans la mesure du possible, la supériorité du médicament expérimental par rapport au traitement de référence. Ces données sont ensuite mises en avant par les équipes d'accès au marché auprès de la CT, afin d'obtenir le meilleur score d'ASMR possible.

L'ASMR peut être qualifiée de majeure (ASMR I), importante (ASMR II), modérée (ASMR III), mineure (ASMR IV) et inexistante (ASMR V).

L'ASMR I correspond à un bouleversement thérapeutique qui sauve ou change la vie des patients atteints d'une maladie grave (par exemple, les vaccins).

Les ASMR II et III représentent une valeur ajoutée importante eu égard à la qualité de la démonstration et à la sévérité de la maladie (par exemple, les médicaments orphelins et les médicaments anticancéreux).

L'ASMR IV valorise un progrès de faible ampleur par rapport à l'existant du fait d'une démonstration et/ou d'une quantité d'effet faible ou non optimale à ce stade de l'évaluation.

La conclusion par la CT d'une ASMR de niveau V peut signifier un manque de valeur ajoutée par rapport à l'existant, mais elle est souvent le fait d'une immaturité des données disponibles, indiquant que la valeur du médicament est encore incertaine. La CT réévaluera alors le médicament ultérieurement, lorsque de nouvelles données seront disponibles.

Une fois l'ASMR déterminé, le CEPS (Comité Économique des Produits de Santé) négocie le prix du médicament avec le laboratoire pharmaceutique. Le prix est fixé en fonction du niveau d'ASMR, du coût des alternatives thérapeutiques et des prix pratiqués dans d'autres pays européens.

Qu'est-ce que l'ASMR ? Quelles sont ses conséquences ?

L'ASMR répond à la question : le médicament apporte-t-il un progrès par rapport aux traitements disponibles ? Si oui, lequel ?

→ Il est apprécié pour chaque indication, par rapport aux alternatives déjà disponibles

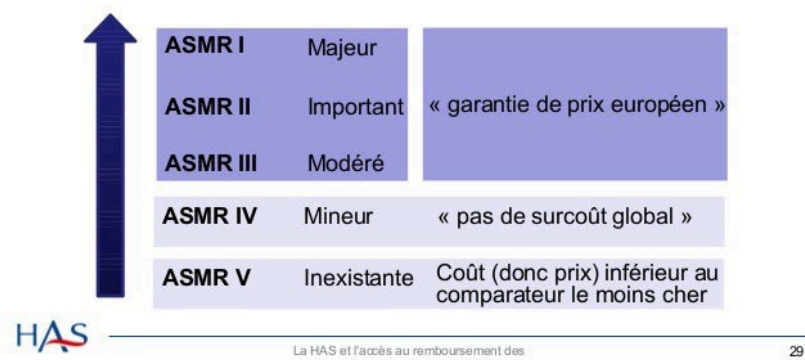


Figure 4. Conséquences de l'ASMR sur la négociation du prix d'un médicament¹⁰

L'obtention d'une AMM, ainsi que l'obtention d'un prix profitable et d'un taux de remboursement intéressant, sont autant de paramètres selon lesquels un médicament pourra ou non voir le jour sur le marché dans un pays ou groupe de pays donné. En effet, l'industrie pharmaceutique se doit de rester profitable pour être pérenne, en garantissant que le chiffre d'affaires réalisé permette de couvrir l'ensemble des frais engagés pour le développement du médicament, ce montant étant estimé à environ à un milliard de dollars.¹¹

Il est donc clair que les autorités de santé jouent un rôle crucial dans l'innovation médicale et on comprend alors qu'il est particulièrement important de prendre en compte les critères qui intéressent ces autorités de santé, afin de maximiser les chances de voir un médicament arriver sur le marché, et être disponible pour les patients qui en ont besoin.

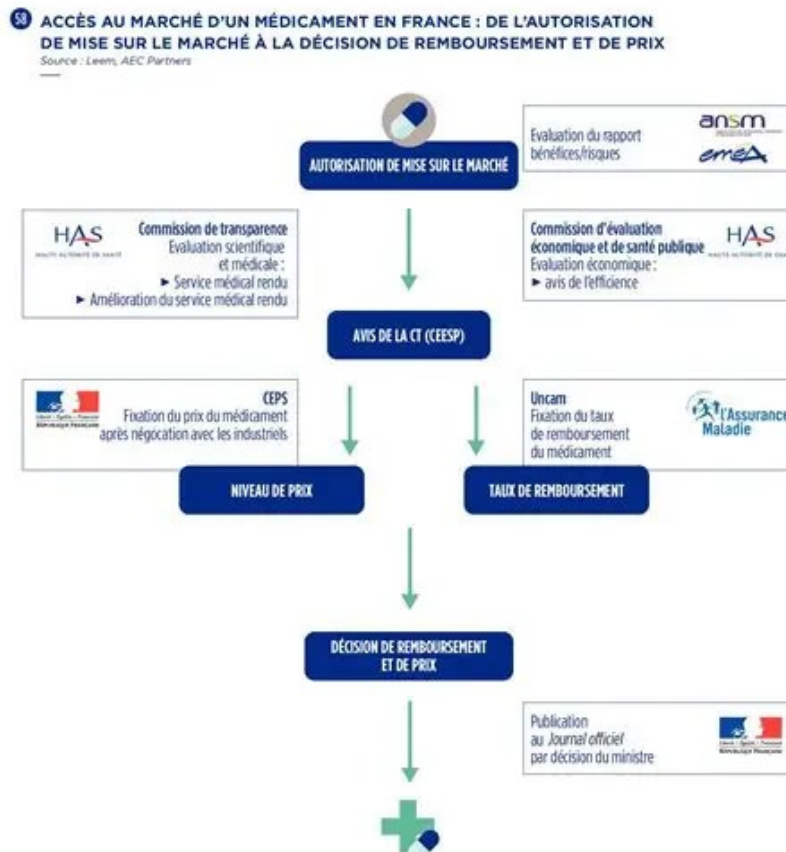


Figure 4. Processus réglementaire de mise sur le marché d'un médicament¹²

2) Le Gold Standard des essais cliniques

Depuis plusieurs décennies, ces autorités de santé plébiscitent un modèle bien précis d'essai clinique, lorsqu'il s'agit de délivrer des autorisations de mise sur le marché et d'octroyer des prix et taux de remboursement intéressants. Ce modèle est celui de l'essai de phase III, contrôlé, randomisé et mené en double aveugle.

Comme nous l'avons vu précédemment, l'essai de phase III est la phase la plus avancée d'un programme d'étude clinique, avant la mise sur le marché. Contrairement aux phases I et II, les essais de phase III incluent généralement un très grand nombre de patients, permettant ainsi d'avoir des données importantes qui soient les plus proches possibles de la vie réelle.

Le terme « contrôlé », signifie que l'essai comporte un ou plusieurs bras dit(s) « expérimental(aux) », dans lequel les patients recevront le traitement qu'on souhaite évaluer, et un bras dit « de contrôle », dans lequel les patients seront traités avec le

traitement de référence actuellement utilisé dans la pratique courante, ou avec un placebo, si aucun traitement n'existe sur le marché.

Le terme « randomisé », fait référence à la méthode d'attribution du bras de traitement pour un patient donné. En effet, lorsqu'un essai est dit randomisé, le bras de traitement sera attribué de façon aléatoire aux patients, par un algorithme prédéfini avant le début de l'étude. Des caractéristiques cliniques spécifiques sont définies en début d'étude, comme pouvant potentiellement influencer sur les effets du traitement. L'algorithme fait alors en sorte de rendre les différents bras de traitement homogènes vis-à-vis de ces caractéristiques (âge, sexe, origine ethnique, comorbidités, traitements concomitants...).

Enfin, lors d'un essai dit « en double aveugle », ni le patient, ni le médecin investigateur ne savent quel traitement a été attribué par l'algorithme. Cela permet de s'affranchir des biais liés à la subjectivité du suivi, pour l'évaluation des critères de jugement des résultats.

Cette méthode permet de rendre les deux groupes étudiés homogènes et comparables, avec comme seule différence le traitement alloué, et donc d'attribuer de manière certaine les différences observées en termes d'efficacité et de sécurité, aux différentes stratégies thérapeutiques étudiées (traitement expérimental ou comparateur/placebo). Cette méthode permet l'imputabilité des résultats au médicament évalué.

C'est ce qu'on appelle le « Gold Standard » en recherche clinique. Ce type d'essai apporte en effet le niveau de preuve le plus important en matière d'évaluation d'un médicament. Ce gold standard est reconnu par les autorités compétentes telles que la FDA (Food and Drugs Administration), l'EMA et la HAS.

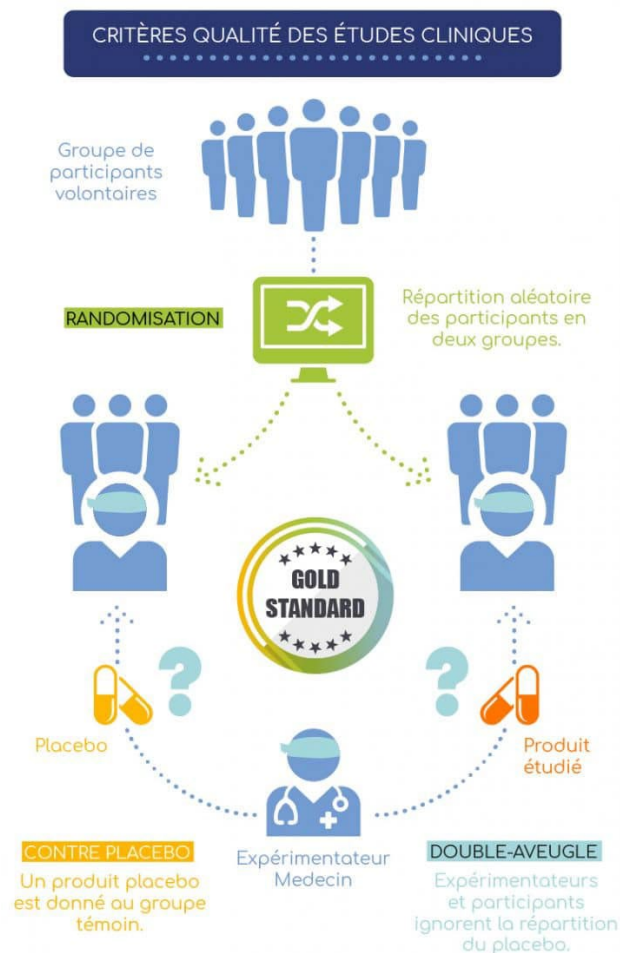


Figure 5. L'essai clinique contrôlé, randomisé et en double aveugle, le Gold Standard qualité d'un protocole d'étude clinique¹³

3) Evolution des objectifs et de la conduite des essais cliniques

Cependant, la recherche clinique est en perpétuelle évolution, et les récentes tendances ne font que le confirmer. En effet, on a longtemps parlé de « blockbuster », terme qui désigne un médicament apportant plus d'un milliard de dollars de chiffre d'affaires. Ce type de médicament n'était pas rare à une époque où tous les patients atteints d'une même maladie étaient traités de la même manière, sans prendre en compte leurs caractéristiques individuelles.

Aujourd'hui, l'industrie pharmaceutique est davantage entrée à l'ère du « custombuster »¹⁴ qui, comme son nom l'indique, est un modèle personnalisé et centré sur le patient et ses caractéristiques cliniques individuelles.

Plusieurs facteurs sont à l'origine de l'évolution des méthodes en matière de recherche clinique.

a. Médecine personnalisée

On peut dans un premier temps citer la médecine personnalisée, qui régit de plus en plus la prise en charge médicale des patients. Cette médecine a pour objectif d'améliorer l'efficacité des traitements, tout en limitant leurs effets indésirables, en ciblant précisément certaines protéines ou voies de signalisation impliquées dans une maladie donnée. Elle a pu voir le jour à la suite des progrès de la médecine, qui permettent jour après jour une meilleure compréhension des processus physiopathologiques à l'origine des maladies. Notamment, l'essor de la génétique permet depuis quelques temps le développement de traitements extrêmement ciblés.

Cette médecine, de plus en plus précise et personnalisée, impose une évaluation plus fine de traitements pour des sous-populations données, compliquant ainsi le recrutement dans les essais cliniques et l'analyse des résultats.

En effet, prenons l'exemple du cancer du poumon, second cancer le plus fréquent chez l'homme et troisième cancer le plus fréquent chez la femme. Cette pathologie dans sa globalité représente une très grande population de patients, avec environ 53 000 nouveaux cas par an en France¹⁵.

Si nous nous intéressons en particulier au cancer du poumon non à petites cellules (CPNPC), celui-ci représente environ 85% de l'ensemble des cancers du poumon¹⁶. La population de patients est légèrement diminuée, mais reste importante.

Si nous nous concentrons maintenant sur le CPNPC causé par la mutation du gène codant pour le récepteur HER2, celui-ci ne concerne que 2% de l'ensemble des CPNPC¹⁷, ce qui réduit désormais considérablement le nombre de patients concernés.

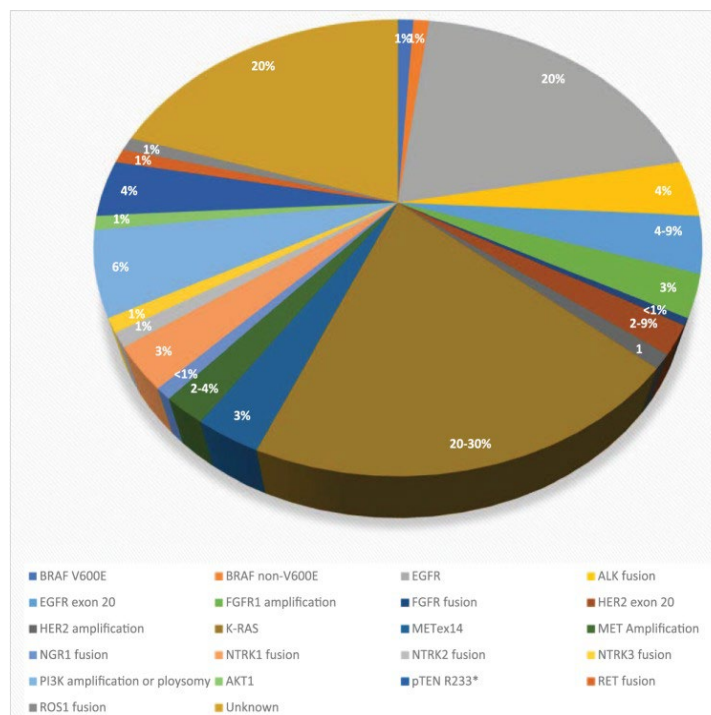


Figure 6. Distribution des différentes mutations activatrices dans le cancer du poumon non à petites cellules chez les populations asiatiques et blanches.¹⁸

La problématique est donc très différente, selon qu'on cherche à développer un traitement contre le CPNPC, ou un traitement ciblant de manière spécifique un sous-type bien précis de ce cancer, ayant en commun la mutation du gène HER2, avec l'objectif d'améliorer l'efficacité du traitement tout en en réduisant les risques.

b. Enjeux de la médecine personnalisée

Les enjeux pour les essais cliniques sont alors multiples, avec en premier lieu un défi important lié à l'identification des patients. En effet, la médecine de précision nécessite une identification précise de sous-groupes de patients, caractérisés par la présence ou l'absence de mutations ou de biomarqueurs spécifiques. L'identification de ces éléments fait appel à des techniques génétiques ou biologiques encore trop peu répandues, voire inexistantes, ce qui peut rendre difficile le recrutement pour les essais cliniques. En oncologie, les biomarqueurs prédictifs ont été associés à environ 60% des nouveaux médicaments oncologiques mis sur le marché¹⁹.

Dans notre exemple de CPNPC avec une mutation du récepteur HER2, il sera nécessaire de mettre en évidence cette mutation avant de pouvoir inclure un patient dans un essai clinique. Or, cette recherche de mutation n'est pas systématiquement

faite d'emblée en pratique courante pour un nouveau patient présentant un cancer du poumon. De plus, 28 plateformes INCA existent en France, ce qui offre un maillage territorial certes important, mais qui peut quand même poser un problème d'accessibilité pour des patients éloignés géographiquement ou socialement du parcours de soin.

Ce n'est qu'en 2021 que l'accès en routine au séquençage génomique devient progressivement pris en charge par l'Assurance maladie pour les patients atteints de certains types de cancers.



Figure 7. Situation des 28 plateformes hospitalières de génétique moléculaire des cancers²⁰

Dans certain cas, il n'existe pas encore de test permettant d'identifier un biomarqueur spécifique, afin de prédire la réponse à un traitement, ou le développement d'évènements secondaires. L'enjeu est alors de développer, en parallèle du traitement, un tel test, appelé test compagnon.

Un autre défi inhérent à la médecine de précision est lié aux données collectées pendant l'essai. En effet, nous l'avons vu, il est difficile d'identifier des patients pour un essai clinique. Cette difficulté est accrue par les critères d'inclusion et d'exclusion des essais cliniques, souvent très restrictifs. Dès lors, un essai clinique de médecine de précision mené selon la méthode traditionnelle peut ne pas fournir suffisamment de données pour évaluer l'efficacité et la sécurité des traitements, à cause du nombre limité de patients inclus. De plus, cela pourrait inclure certains biais, car ces essais sont très ciblés et ne tiennent pas compte de la diversité des populations, ce qui peut affecter la généralisation des résultats.

Ces problématiques risquent ainsi d'augmenter la durée du développement du médicament, car il faut collecter un très grand nombre de données pour satisfaire aux exigences actuelles des autorités de santé.

Enfin, la médecine de précision présente également une problématique relative au comparateur.

Comme nous l'avons dit précédemment, les autorités de santé demandent aujourd'hui la présence d'un bras contrôle, dans lequel les patients reçoivent un placebo ou le traitement de référence, s'il en existe. Cependant, les traitements standards peuvent évoluer rapidement, au gré des découvertes réalisées et des nouveaux médicaments mis sur le marché, rendant difficile la sélection d'un comparateur qui reste pertinent tout au long de l'essai. En effet, lors de l'évaluation d'un médicament par les autorités de santé, celles-ci se basent sur le traitement de référence en vigueur au moment de l'évaluation, qui n'est pas forcément le même que celui choisi lors de la conception du protocole d'étude clinique.

De plus, il est fréquent que les essais cliniques de médecine de précision incluent plusieurs sous-groupes différents de patients qui ont donc des caractéristiques à la fois variées et précises, ce qui complique la sélection d'un comparateur qui soit applicable à l'ensemble des sous-groupes de l'essai.

Également, dans certains cas, il peut ne pas exister de traitement actif approprié existant à utiliser comme comparateur et administrer un placebo pour une partie des patients de l'essai peut dans ce cas poser de réelles questions éthiques. En effet, dans l'exemple d'un essai clinique visant à tester un nouveau médicament anticancéreux, il est inconcevable qu'une partie des patients inclus reçoivent uniquement un placebo.

4) Exemple du développement du Larotrectinib

Prenons l'exemple du Larotrectinib (Vitrakvi®), molécule développée par Loxo Oncology, en partenariat avec Bayer, pour le traitement de cancers très particuliers. Ce traitement cible en effet un évènement génétique bien précis, à l'origine de cancers variés : la fusion des gènes NTRK.

Les fusions NTRK sont retrouvées aussi bien chez les adultes que chez les enfants, dans 45 formes tumorales différentes pour un total de 134 sous-types histologiques. Du point de vue de la prévalence, elles n'interviennent que très rarement dans de nombreux cancers fréquents, comme : le cancer du poumon non à petites cellules

(CPNPC), le cancer du sein, le cancer colorectal, le sarcome des tissus mous, le cancer papillaire de la thyroïde, le mélanome et le glioblastome. Néanmoins, ces réarrangements de gènes apparaissent avec une fréquence plus élevée dans certains types de néoplasmes très rares, tels que : le fibrosarcome infantile (la fusion ETV6-NTRK3 est présente dans 90% des fibrosarcomes infantiles), le mésonéphrome congénital, où la fusion ETV6-NTRK3 est aussi récurrente, le carcinome mammaire sécrétoire et le néoplasme analogue des glandes salivaires²¹.

a. Essais cliniques et résultats

Un essai clinique de phase I a été lancé en 2014 par Bayer pour tester ce médicament chez des patients adultes atteints de tumeurs solides, sans restriction dans le type de cancer, du moment que le patient présentait la fusion recherchée.

Un an plus tard, des premiers résultats d'efficacité clinique du Larotrectinib ont été publiés concernant un premier patient inclus dans cette étude, et pour lequel une fusion du gène NTRK avait été démontrée. Il s'agissait d'une femme de 41 ans atteinte d'un sarcome des tissus mous, métastatique au poumon, ayant reçu 100 mg de Larotrectinib deux fois par jour.

Une régression tumorale clinique rapide a été observée, puisque dès la fin du premier mois de traitement, la patiente était en réponse partielle selon les critères d'évaluation RECIST (Response Evaluation Criteria in Solid Tumors) version 1.1²², avec une nette amélioration de ses métastases pulmonaires. Puis, sur le scanner réalisé 4 mois après le début du traitement, les plus grandes lésions tumorales avaient pratiquement disparu. De plus, la patiente n'a pas eu d'évènement indésirable relié au Larotrectinib, et l'oxygénothérapie n'était plus nécessaire pour le traitement symptomatique de sa dyspnée, symptôme fréquent et particulièrement invalidant chez les patients présentant ce type de cancer.²³

Les résultats finaux de l'étude ont montré que, sur 8 patients inclus présentant la fusion NTRK, 7 ont répondu au traitement par Larotrectinib, dont 2 avec une réponse complète et 5 avec une réponse partielle, et 1 avait une maladie stable.

Titre de l'étude / Phase	<i>A Phase I Study of the Oral TRK Inhibitor Larotrectinib in Adult Patients With Solid Tumors</i>	
ClinicalTrials.gov no.	NCT02122913	
Date "cut-off" *	19 Février 2018	
Population N = 70 patients inclus	67 patients évaluable^{**}	8 patients présentant une fusion du gène NTRK^{***}
Taux de réponse globale**** évalué par les investigateurs	12% (8/67)	88% (7/8) IC95% 47-100 Dont 2 réponses complètes, 5 réponses partielles et une réponse stable
Revue radiologique centralisée indépendante	Taux de réponse globale	100% IC95% 63-100 Dont 2 réponses complètes et 6 réponses partielles
	Durée médiane de suivi	27,0 mois
	Durée médiane d'apparition de la réponse	1,9 mois
	Durée médiane de la survie	Non atteinte
	Localisation de la tumeur (nombre de patients)	Cancer des glandes salivaires (3) GIST (2) Cancer de la thyroïde (1) Sarcome des tissus mous (1) Cancer du poumon (1)

Figure 8. Résultats sur quelques critères d'évaluation secondaires de l'essai clinique de Phase I²⁴

* Date à laquelle les nouvelles données ont cessé d'être prises en compte.

** 2 patients ont été exclus car leur maladie n'était pas mesurable.

*** dont un patient présentant une amplification du gène NTRK1.

**** Comprend les réponses complètes et les réponses partielles selon RECIST version

1.1

Une seconde étude clinique, de phase II, a donc été menée chez des patients adultes atteints de tumeurs exprimant un gène de fusion NTRK²⁵. L'étude NAVIGATE est une étude panier (« basket study », en anglais).

Les essais panier évaluent l'efficacité d'un même médicament dans plusieurs pathologies, en se basant sur le mode d'action et de la cible moléculaire du traitement, plutôt que sur la localisation tumorale ou l'histologie de la pathologie.

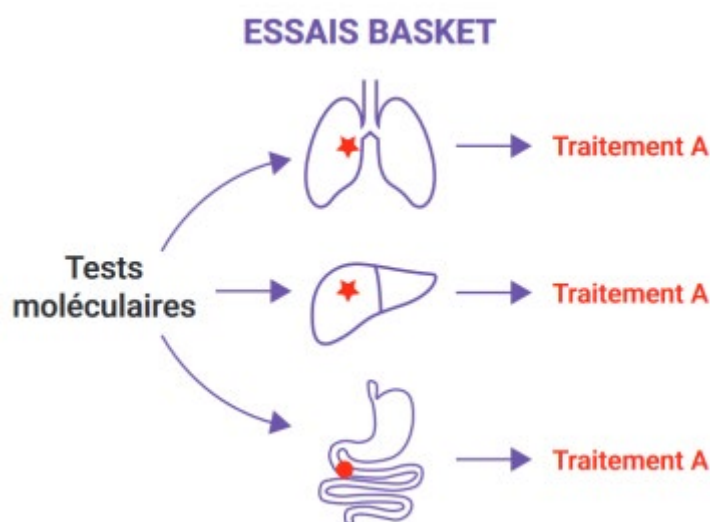


Figure 9. Principe des essais cliniques « basket » : un même médicament peut être étudié pour des cancers de localisations différentes, mais ayant une même anomalie génomique²⁶

Ainsi, l'étude NAVIGATE a testé le Larotrectinib chez des patients atteints de tumeurs différentes, ayant en commun la présence de la fusion du gène NTRK : cancer du poumon non à petites cellules, cancer de la thyroïde, sarcome, cancer colorectal, cancer des glandes salivaires, cancer biliaire, tumeur primitive du système nerveux central (SNC). Une dernière cohorte regroupait les patients présentant une autre histologie que celles citées précédemment.

L'étude SCOUT a quant à elle débuté en décembre 2015 et est toujours en cours. Il s'agit d'un essai de phase I/II portant sur l'inhibiteur de TRK, le Larotrectinib, administré par voie orale, chez des enfants âgés de 1 mois à 21 ans atteints de tumeurs solides avancées ou primitives du système nerveux central. La présence d'une fusion TRK n'était pas requise, sauf pour les nourrissons de moins de 1 an. Cependant, des tests recherchant une fusion du gène NTRK ont été effectués localement avant l'inclusion des patients, ce qui a entraîné un recrutement plus important de patients présentant ce type de fusion par les investigateurs.

Quelques résultats d'efficacité de cette étude sont présentés dans le tableau ci-dessous, démontrant une nouvelle fois une efficacité remarquable du traitement chez les patients présentant une fusion NTRK.

Titre de l'étude / Phase	<i>A Study to Test the Safety and Efficacy of the Drug Larotrectinib for the Treatment of Tumors With NTRK-fusion in Children (SCOUT)</i>	
ClinicalTrials.gov no.	NCT02637687	
Date "cut-off" *	17 Juillet 2017	
Population N = 24 patients inclus	22 patients évaluable^{**}	15 patients présentant une fusion du gène NTRK^{***}
Taux de réponse globale ^{**} évalué par les investigateurs	64% (14/22) IC95% 41-83	93% (14/15) IC95% 68-100 Dont 4 réponses complètes, 10 réponses partielles et une réponse stable
Revue radiologique centralisée indépendante	Taux de réponse globale	93% (14/15) IC95% 68-100 Dont 2 réponses complètes et 12 réponses partielles
	Durée médiane d'apparition de la réponse	1,7 mois
	Durée médiane de la survie	Non atteinte
	Localisation de la tumeur (nombre de patients)	Fibrosarcome infantile (8) Autre sarcome des tissus mous (7)

Figure 10. Résultats sur quelques critères d'évaluation de l'essai clinique de Phase I chez les enfants²⁷

* Date à laquelle les nouvelles données ont cessé d'être prises en compte.

** 2 patients ont été exclus car leur maladie n'était pas mesurable.

*** Comprend les réponses complètes et les réponses partielles selon RECIST version 1.1.

Une première publication des résultats combinés de ces essais cliniques regroupe 55 patients ayant été traités avec le Larotrectinib pour 17 types de tumeurs positives à une fusion NTRK, et étant âgés de 4 mois à 76 ans. Cette analyse groupée de ces essais cliniques faisait partie intégrante du programme de développement du

Larotrectinib depuis le départ, du fait de la rareté des fusions NTRK. D'ailleurs, les premiers dossiers de demande d'AMM se basent sur ces résultats. A la date du 17 juillet 2017, le taux de réponse globale évalué par un comité de revue radiologique indépendant, était de 75% (IC95% 61-85), incluant 7 patients en réponse tumorale complète, et 34 en réponse partielle. Le taux de réponse globale évalué par les investigateurs, était de 80% (IC95% 67-90).

Puis, le 19 février 2019, d'autres résultats poolés ont été publiés comprenant cette fois 159 patients, âgés de 1 mois à 84 ans. 121 patients ont répondu au traitement à l'étude sur les 153 patients évaluable dont 24 réponses complètes observées.

Concernant la tolérance du traitement, l'analyse de sécurité a pris en compte tous les patients (adultes et enfants) ayant reçu le Larotrectinib, sans prendre en compte le statut de fusion NTRK. Les effets secondaires étaient principalement de grade 1 ou 2. Ceux de grade 3 ou 4 reliés au Larotrectinib étaient rares (augmentation de l'alanine aminotransférase, anémie, et diminution des neutrophiles). 5% des patients traités ont eu un événement indésirable grave qui était relié au Larotrectinib (augmentation de l'alanine/aspartate aminotransférase et nausée).

Étant donné la rareté de cette fusion et les difficultés à réaliser un criblage moléculaire de manière systématique, il a été nécessaire d'avoir recours à des approches novatrices, afin d'obtenir un maximum de données cliniques.

Ainsi, le Larotrectinib (Vitrakvi®) est l'un des tout premiers traitements « tissue-agnostic » (qui signifie « indépendant du type tumoral ») approuvé en oncologie : un inhibiteur sélectif des récepteurs TRK indiqué pour des tumeurs solides portant une fusion des gènes NTRK (NTRK1/2/3), indépendamment de l'histologie tumorale.

Cette approche, fondée sur des études « basket » multi-tumeurs, a bousculé les cadres classiques de développement et d'évaluation réglementaire et économique, et a rendu l'accès au marché particulièrement complexe et atypique.

b. Accès au marché

Grâce à une procédure accélérée, la FDA a approuvé, fin 2018, le Larotrectinib en monothérapie chez les adultes et les enfants atteints de tumeurs solides avec fusion NTRK, métastatiques ou pour lesquelles une résection chirurgicale entraînerait une morbidité sévère, en l'absence d'options thérapeutiques satisfaisantes ou après échec thérapeutique, avec une obligation de fournir des données confirmatoires post-AMM.

Au sein de l'Union européenne, l'EMA a délivré une autorisation de mise sur le marché conditionnelle en 2019) pour une indication « agnostique » similaire aux Etats-Unis, couvrant adultes et enfants, assortie d'engagements de génération de données supplémentaires (survie, sécurité à long terme, confirmations par sous-groupes) et d'une demande de mise à disposition d'une formulation pédiatrique (solution orale) afin d'assurer l'accès en pédiatrie.

Dans les autres pays (Canada, Royaume-Uni, Australie, Japon, etc.), des approbations réglementaires généralement alignées sur l'étiquette « agnostique » ont été délivrées.

c. Obstacles et freins à la mise sur le marché

Cependant, des écarts importants ont été observés entre les indications réglementaires incluses dans l'AMM et les indications de remboursement octroyées.

Ainsi, plusieurs autorités d'évaluation des technologies de santé (HTA) ont restreint l'accès remboursé, par exemple via des fonds d'accès conditionnel, des critères stricts de ligne de traitement, ou des segments par types tumoraux, conduisant à un accès réel plus étroit que l'indication « pan-tumeur » annoncée dans l'AMM.

Par exemple, en France, le remboursement n'a été octroyé que dans le cas d'une "impasse thérapeutique", c'est-à-dire si toutes les options de référence adaptées au type tumoral ont été épuisées ou ne sont pas envisageables²⁸. En pratique, cela exclut l'usage précoce et l'usage en alternative à une chirurgie potentiellement réalisable.

De plus, le remboursement n'est autorisé que pour un usage hospitalier avec une décision prise en réunion de concertation pluridisciplinaire (RCP) et avec une confirmation de la fusion NTRK par test validé (souvent NGS, idéalement à partir d'ARN), réalisé dans un laboratoire accrédité.

Plusieurs éléments ont été reprochés par les autorités médico-économiques aux essais cliniques menés pour évaluer le Larotrectinib.

En premier lieu, les critères de jugement choisis pour les essais portant sur le Larotrectinib étaient le taux de réponse objective (ou Overall Response Rate (ORR), en anglais) et la durée de réponse (Duration of Response (DoR), en anglais). Ces critères sont souvent indispensables en développement précoce ou dans des

populations rares, mais ils ne sont pas considérés comme idéaux ou suffisants par la plupart des autorités médico-économiques/HTA pour juger du bénéfice clinique et de la valeur économique d'un traitement. Ce sont en effet des critères qualifiés d'intermédiaires, dont la corrélation avec la survie globale (Overall Survival (OS), en anglais) et la qualité de vie n'est pas universelle ni constante selon les tumeurs. Par ailleurs, l'ORR et la DoR ne disent rien sur le soulagement des symptômes ni la durée chez l'ensemble des patients, la DoR ne concernant que les patients répondeurs, ce qui peut biaiser l'appréciation globale du bénéfice.

L'ORR et la DoR sont privilégiés dans les essais cliniques de maladies rares pour plusieurs raisons. Premièrement, les essais dans les maladies rares incluent peu de patients, il est donc intéressant d'utiliser l'ORR qui est un critère binaire qui nécessite souvent moins de sujets qu'un critère de survie globale (OS) ou de progression (PFS).

Également, dans un essai mono-bras, ce qui peut souvent être le cas dans les maladies rares, on peut tester si l'ORR observée dépasse une valeur historique de référence via un test binomial exact ou une approximation normale.

Un autre avantage est que l'ORR est observable tôt (au bout de quelques semaines ou quelques mois), tandis que la PFS ou l'OS demandent un suivi plus long, souvent impraticable dans des maladies ultra-rares.

Pour des raisons éthiques, les essais contrôlés randomisés sont parfois impossibles lorsqu'il y a trop peu de patients, un besoin médical trop urgent, ou une forte pression éthique à offrir le traitement. Dans ce cas, l'ORR permet des essais mono-bras interprétables avec l'utilisation de comparateurs historiques.

Les instances leur préfèrent des critères dit "patient-relevant" (centrés sur le patient) tels que la survie globale (OS), la qualité de vie rapportée par les patients (HRQoL), les symptômes, les événements cliniques, la survie sans progression (Progression-Free Survival (PFS), en anglais), le temps jusqu'à progression (Time to Progression (TTP), en anglais), ou le temps jusqu'au traitement suivant.

Par exemple, en France, la HAS a tendance à considérer que les ORR/DoR seuls mènent d'office à une ASMR mineure/modérée.

Les HTA exigent donc souvent que des études confirmatoires avec OS/HRQoL ou des comparaisons plus solides soient menées.

Les instances médico-économiques ont également reproché que les données collectées soient issues de cohortes à faible effectif et hétérogènes, sans essais randomisés versus le traitement de référence. En effet, les études « basket » sont

généralement non comparatives car il n'est pas possible de trouver un seul comparateur pertinent pour l'ensemble des cohortes. Chaque cohorte comporte également un faible effectif de patients, du fait de la rareté de la fusion NTRK. Ces conditions sont acceptables pour une autorisation accélérée conditionnelle, mais sont jugées insuffisantes pour un remboursement large et pérenne.

De plus, lors de la modélisation médico-économique inhérente à la négociation des prix et remboursements des médicaments, les HTA établissent un score basé sur le rapport coût-efficacité, appelé score ICER, qui a été évalué très incertain pour l'évaluation du Larotrectinib, du fait de l'absence de comparateurs directs et de l'hétérogénéité tumorale. L'ICER correspond à un rapport coût-efficacité différentiel qui compare une nouvelle stratégie à un comparateur de référence.

Un autre frein mis en avant par les autorités de santé était la nécessité de tests validés, tels que le Séquençage de Nouvelle Génération (NGS) et la crainte de faux positifs ou de faux négatifs, notamment avec l'immunohistochimie pan-TRK. Cela implique donc une certaine exigence de validation inter-laboratoires et de traçabilité, parfois assorties de conditions de remboursement spécifiques.

Ainsi, certaines stratégies de négociation et d'accès au marché ont dû être mises en œuvre par le laboratoire pharmaceutique. Par exemple, des accords de partage de risque ont été signés afin de répartir l'incertitude entre le payeur et le fabricant. De plus, il a été décidé d'un remboursement conditionnel fondé sur les résultats du traitement. Ainsi, un paiement conditionné à la performance, évaluée sur la base des critères de jugement jugés pertinents par les autorités, tels que l'OS et la qualité de vie, a été mis en place.

Par ailleurs, une restriction stricte des patients éligibles au traitement a été décidée, permettant l'accès aux seuls patients sans option standard satisfaisante, avec absence de mutations de résistance connues, et une prescription par des centres experts avec expérience du dépistage NTRK. Ces critères restreignent encore l'accès au marché du Larotrectinib.

Les freins des autorités ont donc porté sur la robustesse et la généralisabilité des données, la qualité du diagnostic et la soutenabilité économique. Ils ont été partiellement levés par des accords de remboursement fondés sur les résultats, des

stratégies de dépistage mieux intégrées, des réductions de prix, et la production de données confirmatoires et en vie réelle.

Le Larotrectinib illustre le paradoxe entre l'ambition des thérapies innovantes et les limites actuelles des cadres d'évaluation du médicament. Si l'indication réglementaire est pionnière et large (adultes et pédiatrie, tumeurs à fusion NTRK), l'accès remboursé a souvent été plus restreint et conditionnel. Cet écart tient principalement à la nature des preuves initiales (essais non randomisés, critères de jugement jugés imparfaits), à la difficulté de généraliser un bénéfice pan-tumoral dans des cadres HTA par cancer, à la balance complexité/prix du dépistage et à des ICER très incertains.

L'expérience du Larotrectinib met donc en évidence la nécessité d'une évolution des cadres de pensée et des outils des autorités de santé. Il ne s'agit pas de « baisser » les exigences, mais de mieux adapter les méthodologies réglementaires et médico-économiques aux thérapies ciblées, aux essais agnostiques et aux technologies de diagnostic de nouvelle génération.

III. Essor du Big Data et de l'Intelligence Artificielle

Par ailleurs, un autre facteur important intervient dans l'évolution des anciennes méthodes en matière de recherche clinique. Il s'agit du Big Data et de l'Intelligence Artificielle. En effet, le développement des bases de données massives et l'intégration de plus en plus importante de l'intelligence artificielle font évoluer les pratiques qui étaient jusqu'ici installées.

Le Big Data en santé est défini comme l'agrégation à grande échelle de données massives et hétérogènes, issues des sources que nous allons détailler par la suite. Comme ces données sont trop importantes pour les outils classiques, on utilise des "data lakes", permettant de stocker de grandes quantités de données brutes à faible coût. On utilise également des systèmes de calcul puissants pour stocker, trier et relier entre elles toutes ces informations. Les données doivent préalablement être transformées pour répondre à des standards communs qui permettent à des sources différentes de se comprendre entre elles, c'est ce qu'on appelle l'interopérabilité des bases.

Le machine learning, consiste à apprendre à un ordinateur à repérer des schémas dans ces données, pour faire des prédictions ou prendre des décisions. En lui montrant des exemples, il apprend à reconnaître ce qui se ressemble (par exemple, des images de tumeurs qui répondent ou non à un traitement) et il généralise cela à de nouveaux cas. On peut lui donner la “bonne réponse” pour qu’il apprenne, c’est ce qu’on appelle l’apprentissage supervisé, mais on peut également le laisser organiser les données par ressemblances, c’est ce qu’on appelle l’apprentissage non supervisé.

Concrètement, pour pouvoir utiliser le Big Data dans la santé, il faut préalablement nettoyer et harmoniser les données, puis “entraîner” le modèle et vérifier que le système ne s’est pas contenté de mémoriser, mais qu’il est capable d’extrapoler. On mesure ensuite sa performance en termes de fiabilité et de précision. Lors de l’utilisation de systèmes de machine learning, il est important de vérifier en continu qu’il reste performant et fiable.

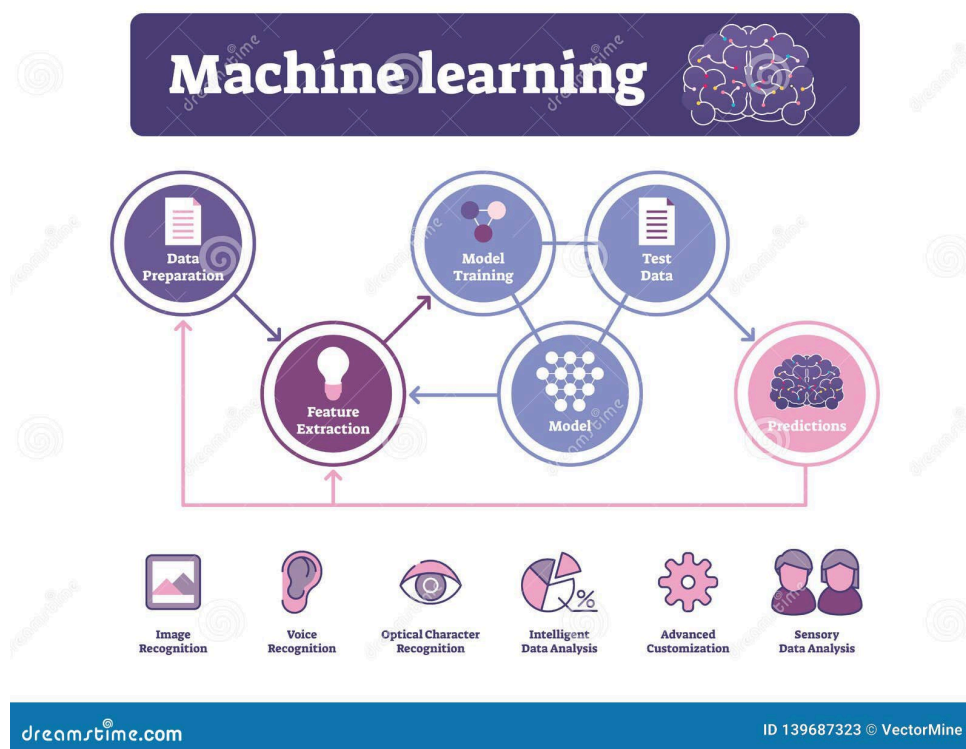


Figure 11. Fonctionnement du machine learning

Le changement de paradigme actuel nous pousse donc à explorer l’impact des évolutions méthodologiques en recherche clinique par rapport au modèle de référence : l’essai de phase III, contrôlé, randomisé, en double aveugle.

Le passage du « blockbuster » au « custombuster », avec des effectifs potentiels de patients plus réduits, et le développement de la médecine personnalisée, sont des éléments d'incitation au recours à ces méthodologies, qui nécessite un plus petit nombre de patients et permettent des délais de développement raccourcis.

L'objectif est donc d'identifier des approches complémentaires à l'essai randomisé en double aveugle, utiles à différents stades du cycle de vie d'un produit de santé, depuis son développement jusqu'à sa mise sur le marché, et son évaluation continue. Ces méthodes visent à maintenir un niveau de preuve élevé tout en répondant aux limites du schéma classique.

Le recours à des designs d'essai clinique alternatifs faisant appel au Big Data et à l'Intelligence Artificielle permettrait une accélération du développement clinique grâce à la disponibilité immédiate de données, ce qui peut donc limiter le temps de collecte et donc les coûts associés, et ainsi permettre une disponibilité plus rapide du traitement sur le marché pour les patients.

De plus, l'utilisation de modèles de simulation tels que les essais in silico permet de réduire le nombre d'inclusions et donc de gagner du temps tout en réduisant les risques potentiellement liés à la participation d'un patient à un essai clinique.

L'organisation de la recherche pharmaceutique doit donc évoluer en profondeur pour relever le défi des nouveaux traitements. Les grands laboratoires s'appuient sur les travaux pionniers d'un nombre croissant de start-ups, et constituent des équipes mixtes avec les centres universitaires de recherche afin de répondre à ces problématiques.

1) Les sources de big data

Les sources de big data utilisables pour les essais cliniques sont variées et peuvent provenir de plusieurs types de données. Nous verrons ci-dessous un aperçu des principales sources, tant en France qu'à l'international.

a. Dossiers médicaux électroniques (DME) et dossiers électroniques du patient (DEP)

Le dossier médical électronique est la version numérique d'un dossier médical papier. Ils contiennent les antécédents médicaux et thérapeutiques du patient dans un établissement de santé donné. Contrairement aux dossiers papier encombrants, les DME permettent un accès rapide et facilité aux données médicales. Les DME renforcent l'efficacité, améliorent la précision des données et offrent des mesures de sécurité solides.²⁹

Le dossier électronique du patient (DEP) est un répertoire numérique des informations médicales d'un patient qui documente l'ensemble de son parcours de soins en temps réel. Diagnostics, antécédents médicaux, procédures, traitements, examens d'imagerie, résultats de laboratoire et notes cliniques sont renseignés et mis à jour électroniquement, de sorte que les médecins et les différents intervenants de soins provenant d'établissements de santé différents disposent des informations les plus récentes lorsqu'ils accèdent au DEP d'un patient.

Ces deux types de dossiers électroniques peuvent être utilisés en tant que source de Big Data. Après analyse, on pourrait ainsi y retrouver de nombreuses informations pertinentes, telles que des signaux et marqueurs spécifiques, ce qui permettrait de pouvoir de diviser les patients en des sous-groupes particuliers.

Par exemple, une étude rétrospective a été menée à partir de données issues de dossiers de santé électroniques, chez des patients atteints de cancer bronchique non à petites cellules métastatique (NSCLC) ayant reçu du pembrolizumab en pratique réelle³⁰. L'objectif de l'étude était d'obtenir des informations sur les résultats cliniques à long terme présentés par ces patients.

Grâce à l'étude de ces sources, on pourrait en effet identifier rapidement et facilement des caractéristiques permettant de prédire l'efficacité ou la sécurité d'un traitement sur l'un ou l'autre de ces sous-groupes. Cela permettrait par exemple de ne pas sélectionner un sous-groupe particulier de patients dont on sait qu'ils ne répondront pas efficacement au traitement, leur évitant ainsi un traitement inutile avec les événements indésirables associés et l'espoir d'une guérison, et permettant de réduire le coût associé à l'inclusion d'un patient dans un essai.

b. Bases de données de santé publique

Il existe de nombreuses bases de données de santé publique exploitables en France et sur le plan international.

En France, on peut citer le Système National des Données de Santé (SNDS)³¹ qui regroupe des données de l'Assurance Maladie et des dossiers médicaux des hôpitaux. L'objectif du SNDS est de faire de la France l'un des pays pionniers dans le domaine de la promotion et de la valorisation des données de santé.

Géré par la [Plateforme des données de santé](#) (PDS ou « Health data hub ») et la [Caisse nationale de l'assurance maladie](#) (CNAM), le SNDS est alimenté, depuis 2016, par la base de données « SNIIRAM » (Système national d'information inter-régimes de l'assurance maladie) contenant les données relatives à toutes les dépenses de l'assurance maladie, la base de données « PMSI » (Programme de médicalisation des systèmes d'information) contenant les données d'analyse de l'activité des établissements de santé, la base de données du CépiDc (Centre d'épidémiologie sur les causes médicales de décès) contenant les données relatives aux causes de décès.

Progressivement et sur le fondement de la loi du 24 juillet 2019, le SNDS sera également alimenté par des données médico-sociales liées au handicap, fournies par les maisons départementales des personnes handicapées.

A l'international, on peut citer le National Health Service (NHS) au Royaume-Uni, ou Medicare aux États-Unis.

Ces bases de données permettent d'analyser des cohortes de patients sur de grandes échelles et de suivre les résultats de santé à long terme.

c. Essais cliniques précédents et registres

A la fin de l'année 2025, on trouve plus de 550 000 essais cliniques, terminés ou en cours, enregistrés sur la base de données internationale des essais cliniques *ClinicalTrials.gov*. Ce chiffre est en perpétuelle augmentation, car de nouveaux essais sont constamment menés. Lors de la conduite d'un essai clinique, les données produites sont strictement confidentielles, mais une fois les résultats collectés et

analysés, le promoteur de l'essai a l'obligation de les publier, qu'ils soient positifs ou négatifs. Ainsi, comment ne pas envisager d'utiliser les données déjà collectées dans des essais cliniques précédents, afin d'optimiser les essais futurs et actuels ?

On peut effectivement imaginer l'utilisation de techniques d'analyse de données avancées, comme le machine learning et l'analyse statistique, pour identifier des tendances et des corrélations dans les données telles que la mise en évidence de facteurs de risques, l'identification de biomarqueurs ou encore la segmentation des populations.

d. Données génomiques

On remarque aujourd'hui un essor de l'analyse du génome humain. En effet, les avancées technologiques permettent désormais de générer, grâce aux techniques de séquençage à haut débit, des données génomiques à partir d'échantillons biologiques de manière plus rapide et moins coûteuse, facilitant ainsi la collecte de ces données à grande échelle.

Ces données génomiques proviennent de diverses sources telles que des bases de données publiques, parmi lesquelles nous pouvons citer « *The Cancer Genome Atlas (TCGA)* », « *GenBank* », « *European Genome-phenome Archive (EGA)* » ou encore « *dbGaP (Database of Genotypes and Phenotypes)* ».

Il existe également des projets de séquençage à grande échelle, tel que le projet « *1000 Genomes* »³² ayant pour but de fournir des données génomiques qui peuvent être utilisées pour la recherche.

De plus, les biobanques collectent et conservent des échantillons biologiques (comme le sang ou les tissus) et des données associées (comme des informations cliniques et démographiques) pouvant être utilisées dans le futur.

Ces données génomiques sont essentielles pour comprendre comment les variations génétiques influencent le développement des maladies et la réponse aux traitements.

En France, les données génétiques sont considérées comme sensibles et exigent donc une base légale stricte, des finalités déterminées, la minimisation de leur

utilisation. Elles sont ainsi gouvernées par le Règlement Général sur la Protection des Données (RGPD) et les comités d'éthique, et doivent répondre à des normes de sécurité renforcée via la pseudonymisation, le chiffrement et un contrôle d'accès stricte.

e. Données de surveillance et de pharmacovigilance

Les données de surveillance et de pharmacovigilance sont issues de déclarations spontanées d'effets indésirables par les patients ou les professionnels de santé, collectées par des organismes de réglementation et des systèmes de pharmacovigilance, et regroupées sur des bases comme VigiBase/EudraVigilance, ou FDA Sentinel.

Elles sont nécessaires à l'évaluation de la sécurité des médicaments pendant leur développement et après leur mise sur le marché. À grande échelle et quasiment en temps réel, elles permettent la détection de signaux, la caractérisation d'événements rares ou retardés, l'identification d'interactions médicamenteuses et l'exploration de risques dans des sous-populations telles que les enfants, les personnes âgées ou les patients avec des comorbidités.

Utilisées en recherche clinique, elles enrichissent la génération d'hypothèses, influent sur la conception d'essais par le choix de critères d'évaluation de la sécurité, l'identification de populations à risque, ou la mise en place d'une surveillance particulière d'un événement spécifique. Bien utilisées et exploitées, ces données transforment la pharmacovigilance en un outil permettant d'enrichir la recherche clinique et la tout en nécessitant validation et prudence dans l'interprétation.

Toutefois, elles sont exposées à des biais majeurs tels que la sous-déclaration, la duplication, la confusion par indication et les erreurs de codage.

f. Données des réseaux sociaux et forums de patients

Les données issues des réseaux sociaux et des forums de patients constituent une source émergente de big data en santé, capturant quasiment en temps réel l'expérience vécue des patients. Elles peuvent refléter symptômes, effets indésirables,

observance, qualité de vie, parcours de soins et besoins non satisfaits, via des récits libres, des publications et des commentaires.

Pour la recherche clinique, elles peuvent permettre de détecter des signaux de sécurité précoces, de repérer des événements rares et de suivre des tendances de symptômes après la mise sur le marché d'un médicament.

Elles aident aussi à concevoir des essais plus centrés sur le patient, en adaptant le choix des critères de jugement ou des critères d'inclusion afin qu'ils soient aussi pertinents que possible pour les patients. Elles peuvent également optimiser le recrutement, en ciblant la localisation des patients ou de communautés actives.

Bien entendu, les contraintes juridiques et éthiques pour l'utilisation de ces données sont majeures, et incluent le respect des Conditions Générales d'Utilisation des plateformes, le respect du RGPD et des bases légales, l'anonymisation/la pseudonymisation des données et la prévention de la ré-identification.

La transparence et l'engagement avec les communautés sont essentiels, ainsi que la gouvernance via les comités d'éthique et le Data Privacy Officer (DPO).

En pratique, ces sources sont complémentaires aux données cliniques formelles, elles enrichissent la recherche clinique, mais requièrent validation méthodologique et prudence interprétative.

g. Données de capteurs et de dispositifs médicaux

Les données de capteurs et de dispositifs médicaux tels que les glucomètres, tensiomètres, oxymètres, électrocardiogrammes (ECG), pacemakers, pompes à insuline connectés, offrent des mesures continues et objectives de plusieurs paramètres en conditions réelles.

Fréquence cardiaque, activité physique, sommeil, glycémie, pression artérielle, taux de saturation en oxygène (SpO₂), respiration, posture, adhérence thérapeutique... sont autant de données pouvant être collectées via ces capteurs et dispositifs médicaux.

En recherche clinique, elles peuvent être utilisées comme des marqueurs numériques, et permettre une surveillance à distance dans le cadre d'essais

décentralisés, avec une détection précoce d'effets indésirables, ou de paramètres influant sur l'état de santé des patients.

Elles améliorent la personnalisation des soins, tout en réduisant les visites à l'hôpital, ce qui influe positivement sur la rétention des patients dans les essais cliniques, qui est primordiale.

h. Données issues de la télémédecine

La télémédecine s'est beaucoup développée suite à la pandémie de Covid-19, rendant plus faciles et sécurisées ces consultations médicales à distance.

Les données issues de la télémédecine par le biais de téléconsultations audio ou vidéo, de chats en ligne médicaux ou d'e-prescriptions forment un flux massif de données exploitables en big data. Elles captent des données similaires à celles collectées par les consultations classiques, à savoir les symptômes, les parcours de soins des patients, les traitements, les comorbidités, la démographie, et bien d'autres encore.

En recherche clinique, elles peuvent faciliter la mise en place d'essais décentralisés, permettre le suivi à distance du patient et la collecte de données à distance.

L'intégration de ces différentes sources de big data dans les essais cliniques pourrait offrir des opportunités considérables pour améliorer la conception, l'efficacité et la sécurité des études cliniques. L'utilisation de ces données pourrait donner aux chercheurs l'opportunité de mieux comprendre les populations de patients, d'affiner les protocoles d'essai et, finalement, d'améliorer les résultats de ces essais cliniques, permettant ainsi la mise sur le marché de nouveaux traitements de manière rapide et efficace.

IV. Utilisation du Big Data dans les essais cliniques

Le secteur de la santé connaît une accélération de l'utilisation des données massives et de l'intelligence artificielle (IA). Les partenariats entre spécialistes de l'IA et grands laboratoires pharmaceutiques se multiplient. On peut par exemple citer le partenariat entre Bristol Myers Squibb et Exscientia³³, ou entre Sanofi et Aqemia³⁴, avec l'appui des géants du numérique tels que Microsoft et Google.

Cette dynamique transforme la pratique du secteur de la Recherche & Développement (R&D) en raccourcissant les cycles de génération d'hypothèses, de criblage, d'optimisation et de validation.

1) Accélération des découvertes

Certains exemples illustrent cette accélération. On peut d'abord citer la société Exscientia qui a co-conçu le DSP-1181, un candidat-médicament pour le traitement des troubles obsessionnels compulsifs (TOC)³⁵. Ce traitement a été présenté en 2020 comme l'un des premiers composés "conçus par IA" à entrer en essai clinique de phase I, car cette molécule a été trouvée en explorant des millions de molécules par deep learning, marquant une étape symbolique vers une découverte plus rapide et ciblée. En effet, pour le développement du DSP-1181, moins de 12 mois se sont écoulés entre le criblage initial et la fin des essais précliniques, tandis que la moyenne mondiale dans l'industrie pharmaceutique est de 4 à 6 ans.

Par ailleurs, au Massachusetts Institute of Technology (MIT), l'identification de l'antibiotique Halicin³⁶ a montré qu'un algorithme bien entraîné sur des bibliothèques de composés pouvait révéler des mécanismes d'action inédits et des familles thérapeutiques inattendues.

Ici, le phénomène majeur permis par l'utilisation du Big Data, est la compression des délais : le criblage "in silico", c'est-à-dire à partir de bases de données, faisant passer l'étape de criblage de plusieurs semaines à quelques jours, voire à quelques heures selon les ressources et le type de tâche, ce qui permet d'allouer plus tôt du temps expérimental aux candidats prometteurs.

Les boucles de machine learning font le lien entre calcul et expérience. Ainsi, un modèle propose dans un premier temps certaines variantes, puis des robots de laboratoire synthétisent et testent ces hypothèses, et les résultats obtenus réalimentent le modèle continuellement, afin d'améliorer itérativement la qualité des prédictions et de prioriser les essais les plus informatifs.

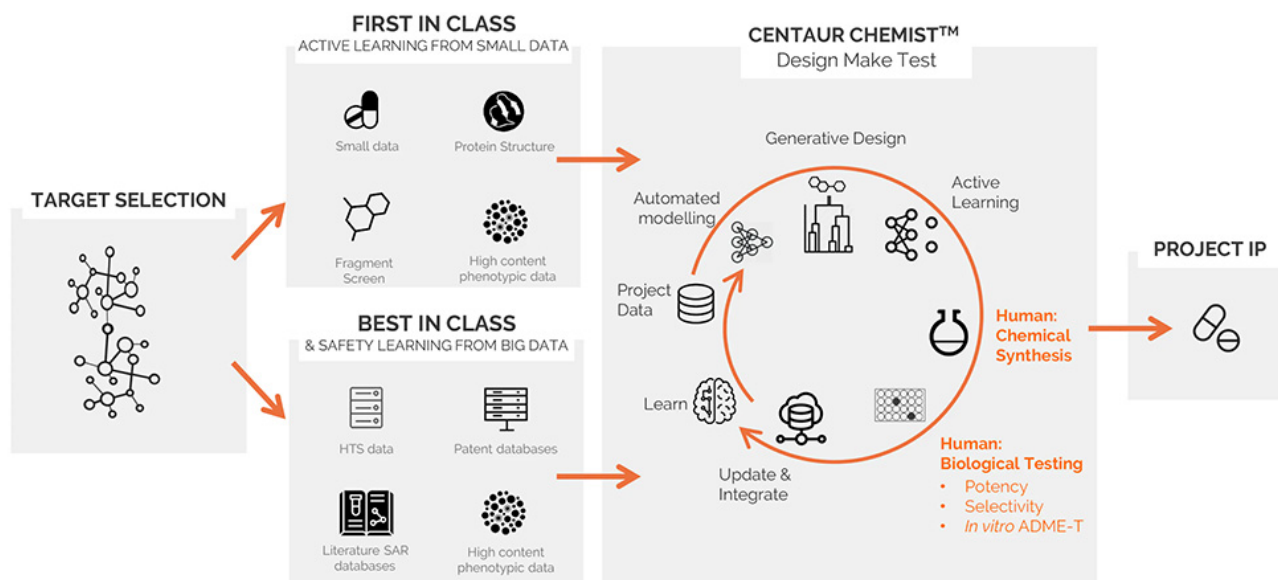


Figure 12. Découverte de candidats-médicaments par machine learning³⁷

Il convient cependant de ne pas surinterpréter le rôle de l'IA dans les découvertes. Elle ne remplace ni la créativité scientifique ni l'expertise biologique et clinique. Ainsi, le criblage "in silico" accélère la formulation d'hypothèses, mais n'offre pas à lui seul des solutions thérapeutiques clés en main.

Le rôle de l'IA est d'abord d'orienter efficacement la recherche, de réduire les ressources ainsi que l'espace et le temps de recherche, de hiérarchiser les pistes, puis de déléguer aux expérimentations la confirmation des mécanismes, de l'efficacité et de la sécurité.

L'accélération la plus féconde est celle issue de la collaboration entre data scientists, chimistes, biologistes et cliniciens, articulée autour d'un protocole de validation rigoureux.

Un autre exemple est celui de l'entreprise Recursion qui se veut être une "fabrique à découvertes"³⁸. En associant des plateformes logicielles d'IA et de la robotique cellulaire, la start-up construit des modèles génétiques de maladies, notamment rares, et génère des bibliothèques d'images à haut contenu afin de

détecter automatiquement des phénotypes d'intérêt à grande échelle. Des centaines de milliers de coupes histologiques sont ainsi analysées pour relier signatures cellulaires et voies thérapeutiques, avec l'objectif de multiplier les candidats testés afin d'identifier rapidement les impasses et de concentrer les ressources sur les hypothèses prometteuses.

Au-delà de la découverte de nouvelles entités, le big data accélère aussi la redécouverte et le repositionnement.

La revue de la littérature et des données cliniques en vie réelle - notamment via l'analyse des différentes sources de Big Data que nous avons détaillées plus haut - permettent d'identifier des usages de médicaments hors-AMM - c'est-à-dire en dehors de l'indication prévue par l'AMM - plausibles et intéressants, des combinaisons synergiques ou des nouvelles indications pour des molécules connues.

Ceci est particulièrement utile dans des domaines où le besoin est urgent, comme dans le domaine des anti-infectieux, des maladies rares ou de l'oncologie de précision, et où les délais de développement sont critiques.

De plus, l'IA peut aider à anticiper les risques (toxicités, interactions) via des modèles prédictifs, réduisant les échecs tardifs très coûteux.

En somme, l'“accélération des découvertes” grâce à l'IA désigne la capacité à transformer plus vite des masses de données en hypothèses testables, puis en candidats validés, grâce à une boucle intégrée calculs - tests au laboratoire. Les partenariats industriels et académiques montrent que l'on peut réduire le temps de criblage, améliorer la sélection et diminuer les risques au plus tôt, sans sacrifier la rigueur scientifique.

Dans ce domaine, l'avenir le plus fécond n'oppose pas l'IA et l'expertise humaine : il les combine pour explorer intelligemment, expérimenter rapidement et décider mieux, afin que la vitesse serve la qualité et, au final, des bénéfices tangibles pour les patients.

2) Optimisation de la conception des essais

Le big data pourrait jouer un rôle crucial dans l'optimisation de la conception des essais cliniques, permettant d'obtenir des résultats probants plus rapidement et de manière fiable. Voici quelques façons dont le big data peut contribuer à cet objectif.

Dans un premier temps, on pourrait imaginer l'utilisation du big data pour l'identification des patients potentiels répondeurs au traitement expérimental. Grâce à l'analyse de grandes quantités de données provenant de précédents essais cliniques, de dossiers médicaux électroniques et d'études génomiques, il est possible d'identifier des biomarqueurs et des caractéristiques démographiques qui prédisent la réponse à un traitement expérimental. Cela permet de sélectionner des patients qui sont plus susceptibles de bénéficier du traitement, améliorant ainsi l'efficacité de l'essai, tout en évitant à une partie de la population de tester un médicament qui a peu de chance de fonctionner sur elle.

Par ailleurs, le Big Data peut également être utilisé dans la sélection des sites d'investigation, afin d'optimiser leur implantation pour améliorer le recrutement. En effet, le Big Data peut permettre de générer une cartographie des prévalences des pathologies ciblées, de comprendre les parcours de soins des patients cibles, et de pré-identifier des patients potentiellement éligibles.

Aussi, le big data peut être utilisé pour stratifier les patients au sein des essais. La stratification des patients dans un essai clinique est un processus qui consiste à diviser les participants en sous-groupes basés sur des caractéristiques spécifiques avant le début de l'essai. Ces caractéristiques peuvent inclure des facteurs démographiques (âge, sexe), des critères cliniques (stade de la maladie, comorbidités), des biomarqueurs, ou d'autres variables pertinentes qui pourraient influencer la réponse au traitement. Les résultats de l'étude sont alors présentés en fonction des différents sous-groupes, permettant d'augmenter la probabilité de démontrer une efficacité du traitement sur un ou plusieurs sous-groupes.

Le big data permet, dans un premier temps, d'identifier des caractéristiques pertinentes pour la stratification. Par exemple, en analysant une grande quantité de données de santé, on peut se rendre compte de l'existence de certains biomarqueurs prédictifs de la réponse à un traitement ou à une classe de médicaments. Les modèles prédictifs basés sur le big data peuvent estimer les résultats potentiels pour différents

sous-groupes de patients. Cela aide à affiner la stratification en se concentrant sur les groupes les plus susceptibles de bénéficier du traitement.

L'analyse des données historiques peut également révéler des tendances dans des résultats antérieurs, permettant aux chercheurs de concevoir des protocoles d'essai plus efficaces, par exemple, en ajustant les doses ou en modifiant la durée des traitements en fonction des données collectées précédemment.

Par ailleurs, l'utilisation de technologies de big data permet de surveiller les données des essais cliniques en temps réel. Cela facilite l'identification rapide des effets secondaires ou des réponses au traitement, permettant des ajustements immédiats si nécessaire par le biais d'amendements au protocole.

En optimisant le processus de recrutement et en améliorant la conception des essais, le big data peut réduire le temps et les coûts associés à la réalisation d'essais cliniques, ce qui est particulièrement important dans le développement rapide de nouveaux traitements.

En résumé, le big data permet d'affiner la conception des essais cliniques en se basant sur des analyses approfondies des données, ce qui conduit à des résultats plus probants, une identification plus précise des répondants au traitement, et une exécution plus rapide et fiable des essais.

Un exemple particulièrement intéressant de l'utilisation du Big Data pour la conception et la conduite d'un essai clinique est celui de l'étude RECOVERY, menée au Royaume-Uni dans le cadre de la recherche de traitement pour la Covid-19, en 2020³⁹.

Une fois que le patient avait consenti à participer à l'essai, seules quelques données essentielles étaient à collecter au lit du patient : identité, âge, comorbidités, niveau d'oxygène... Pas de longs formulaires à remplir ni de visites multiples dédiées à l'essai à réaliser.

Pour mesurer les résultats (décès, durée d'hospitalisation, passage en soins intensifs, recours à l'assistance respiratoire, réadmissions), l'essai n'exigeait pas que le patient revienne à intervalle régulier à l'hôpital, ni que chaque hôpital renvoie manuellement toutes ces informations dans une base de données. À la place, les identifiants des participants tels que le numéro national de santé ou la date de

naissance ont été utilisés pour faire correspondre automatiquement les patients de l'essai avec les grandes bases du système de santé : registre national des décès, bases d'admissions et de séjours hospitaliers, bases des tests de laboratoire, bases des soins intensifs...

À intervalles réguliers, le centre de coordination recevait des extractions de ces bases et devait les raccorder au fichier des participants, de manière pseudonymisée et sécurisée.

Cette approche a permis d'avoir un suivi quasi automatisé et ultra simplifié, tout en étant très complet et homogène, même quand des patients passaient d'un hôpital à un autre. L'un des avantages est de réduire fortement les délais et le travail des équipes de recherche locales, ainsi que les coûts, tout en améliorant la qualité des données et en préservant au maximum le confort des patients.

L'intérêt dans le contexte de la pandémie était important car, comme les informations arrivent vite et à grande échelle, l'essai a pu analyser les résultats en continu et adapter rapidement ses conclusions dans l'objectif d'identifier un traitement bénéfique ou inefficace en quelques semaines, ce qui était particulièrement utile dans le contexte d'urgence à cette période.

Nous pouvons également citer l'essai ADAPTABLE⁴⁰, mené aux Etats-Unis, qui est un vaste essai comparant différentes doses d'aspirine. Dans cet essai, les participants ont été identifiés directement dans les dossiers médicaux électroniques via l'intelligence artificielle, puis ont été inclus via un consentement en ligne. L'essai a ensuite suivi l'évolution de ces patients via les données de routine, ce qui a réduit les coûts, les délais, et l'impact sur les patients.

3) Essais clinique innovants

a. Cohortes synthétiques, ou cohortes externes

Les cohortes synthétiques sont des ensembles virtuels de patients créés à partir de données réelles, grâce à des techniques avancées de modélisation et de simulation informatique s'appuyant sur le Big Data et l'intelligence artificielle. Elles visent à reproduire fidèlement les caractéristiques démographiques, cliniques et biologiques d'une population cible, permettant ainsi de prédire les résultats des traitements médicaux et de compléter les études cliniques traditionnelles.

L'origine des cohortes synthétiques remonte aux années 1980 et 1990, avec l'émergence de l'informatique biomédicale et les premières tentatives de modélisation épidémiologique pour comprendre la propagation des maladies infectieuses. Au début des années 2000, l'augmentation des capacités de calcul et la disponibilité croissante des données de santé électroniques ont permis aux cohortes synthétiques de gagner en importance dans la recherche biomédicale et pharmaceutique. Le développement des algorithmes d'intelligence artificielle (IA) et d'apprentissage automatique conduit aujourd'hui à la création de modèles de plus en plus précis et représentatifs des populations humaines.

L'intérêt principal de ces cohortes synthétiques intervient lorsqu'il existe de fortes probabilités que le bras expérimental soit supérieur au bras standard. Il est alors souhaitable de minimiser au maximum l'exposition de patients au traitement standard. Les méthodologies hybrides, combinant des patients réels et des données de vie réelle, permettent ainsi de réduire le nombre de patients exposés à des traitements potentiellement moins efficaces.

Dans d'autres situations, l'essai clinique comparatif et randomisé idéal se heurte à des contraintes majeures : effectifs de patients trop faibles dans les maladies très rares, enjeux éthiques lorsqu'un bras contrôle implique un placebo, enjeux des essais pédiatriques... Dans ces différents contextes, constituer un bras témoin de taille suffisante peut relever de la « mission impossible »⁴¹.

Les simulations virtuelles que constituent les cohortes synthétiques s'appuient sur des modèles pharmacocinétiques (PK), pharmacodynamiques (PD), et pharmacologiques. Elles permettent de prédire l'association entre l'exposition au traitement et la réponse, en tenant compte de variables pertinentes telles que l'âge, la fonction rénale, les comorbidités, les biomarqueurs exprimés par le patient, de la dynamique de la maladie et des interactions traitement-cible.

Les cohortes synthétiques, parfois appelées contrôles externes, composent un bras témoin à partir d'observations réelles extraites de sources de vie réelle harmonisées et alignées sur les critères du protocole, afin de servir de comparateur à un bras traité dans un essai mono-bras ou à contrôle réduit.

Nous pouvons citer l'exemple du Selinexor, inhibiteur sélectif de l'exportine-1, développé dans le traitement du myélome multiple triple-réfractaire chez des patients

ayant épuisé toutes les options thérapeutiques standards⁴². Dans cette population particulière de patients, plusieurs défis rendaient particulièrement compliquée la mise en place d'un essai clinique « standard ». D'une part, car l'espérance de vie médiane sans traitement pour ces patients est de 6 mois. D'autre part, le recrutement pour un essai clinique est extrêmement difficile car les patients sont rares, épuisés par les lignes précédentes, et il est éthiquement discutable d'inclure un bras placebo ou standard sans accès à l'innovation. Enfin, l'historique des traitements pris par ces patients est très hétérogène, rendant la comparaison directe complexe. Le laboratoire promoteur de l'essai clinique, Karyopharm Therapeutics, a alors décidé de lancer l'essai clinique STORM, avec un bras unique de patients, ainsi qu'un bras de contrôle synthétique. Les données ayant permis de construire ce bras synthétique étaient issues de différentes sources, telles que des registres sur le myélome multiple, des données d'essais cliniques précédents et des dossiers médicaux électroniques de patients. Le big data a également été utilisé pour stratifier les patients de manière optimisée, selon leur statut cytogénétique, le nombre de lignes antérieures reçues, ou le type de réfractarité).

La FDA a approuvé le Selinexor en juillet 2019 pour le myélome multiple triple-réfractaire, en acceptant les preuves issues de l'essai à bras unique complétée par les comparaisons externes — sous condition de réaliser une étude post-approbation confirmatoire via l'essai STORM-2 qui était, lui, randomisé.

Toutefois, la validation des modèles et l'acceptabilité réglementaire exigent une démarche précise et systématique. D'une part, la calibration interne permet de vérifier que les distributions simulées reproduisent les profils observés. D'autre part, la validation externe teste la capacité prédictive sur des cohortes indépendantes, et des analyses de transportabilité évaluent la stabilité par site, sous-groupe et période.

Les autorités de santé (FDA, EMA) encouragent le développement informé par les modèles (Model-Informed Drug Development, en anglais). Dans ce type de modèle, plutôt que de s'appuyer uniquement sur des essais successifs, on intègre les connaissances précliniques et cliniques dans des modèles mathématiques afin de prévoir, comparer des scénarios et optimiser le plan de développement.

Cependant, les autorités requièrent également la transparence des méthodes, la pré-spécification des règles d'intégration statistique, ainsi que des audits de qualité.

Les gains de temps offerts par les bras synthétiques doivent toutefois être mis en balance avec la généralisabilité des résultats. En effet, un contrôle virtuel très ciblé sur une sous-population peut améliorer la puissance mais limiter la transportabilité des conclusions, d'où l'intérêt de conceptions hybrides combinant contrôle réel (même réduit) et contrôle virtuel, et de sous-analyses pré-spécifiées.

Un autre exemple d'utilisation des cohortes synthétiques est celui d'essais dans la dystrophie musculaire de Duchenne, maladie rare et particulièrement grave. Dans ce type de pathologie rares, recruter des patients pour des essais cliniques est un vrai défi. De plus, assigner des patients à des groupes placebo sur plusieurs années soulève des questions éthiques et de maintien des participants dans l'étude. Un essai randomisé plus flexible a donc été mené, avec un design innovant permettant de tester plusieurs doses d'un médicament et d'ajuster l'attribution des traitements au cours de l'essai, en fonction de la réponse initiale des participants. Les données des patients sous placebo ont également été complétées par des données déjà disponibles dans un registre d'histoire naturelle de la maladie, pour limiter le nombre de patients dans le bras contrôle⁴³.

b. Bras *in silico*

Les essais cliniques *in silico* ou intégrant un « bras virtuel » permettent de simuler l'effet d'un médicament (ou d'un placebo) sur un groupe de patients virtuels. Ils utilisent des modèles mathématiques appliqués à des données recueillies en vie réelle (registres, cohortes, objets connectés...).⁴⁴ Les bras *in silico* font référence à l'utilisation de modèles informatiques et de simulations imitant les caractéristiques de vraies populations ou groupes d'individus, pour simuler des bras comparatifs dans des études cliniques. Cela permet d'évaluer l'efficacité et la sécurité d'un traitement sans avoir besoin d'un groupe de participants réels, ou en complément des essais cliniques traditionnels.

Les bras *in silico*, tout comme les cohortes synthétiques, pourraient également permettre de réduire la durée d'un essai clinique, car le nombre de patients à inclure serait considérablement réduit. En effet, la durée d'une étude clinique est en partie régie par la durée du recrutement des patients, elle-même fonction du nombre de patients à inclure et de la facilité à trouver et inclure des patients dans l'étude.

Le renforcement de la robustesse des résultats est un autre avantage des bras *in silico* et des cohortes synthétiques. En effet, en utilisant un bras virtuel en remplacement du groupe placebo, on évite ainsi le phénomène bien connu de l'« effet placebo ». L'effet placebo est le phénomène selon lequel un patient peut ressentir des bénéfices cliniques après l'administration d'un traitement qui n'a pourtant aucune activité thérapeutique. Les recherches ont montré que leur administration déclenche des modifications de l'activité cérébrale qui conduisent à la production de molécules favorables à notre sentiment de bien-être, comme des endorphines ou de la dopamine. Les travaux du CNRS montrent également que l'effet placebo ne perd rien de son efficacité si le patient est au courant que c'est un placebo qui lui est administré.⁴⁵

Le remplacement des groupes placebo par des bras *in silico* pourrait donc permettre, en supprimant l'effet placebo, d'accroître la différence d'effet entre le bras placebo et le bras recevant le traitement expérimental, mettant ainsi davantage en valeur la significativité de l'effet du traitement à l'étude.

Un exemple de l'utilisation de bras *in silico* est l'essai virtuel VICTRE mené par l'autorité américaine des dispositifs médicaux, au cours duquel des seins numériques réalistes ont été simulés pour comparer deux technologies d'imagerie mammaire, la mammographie numérique et la tomosynthèse mammaire numérique. Cet essai a été mené entièrement *in silico* avec 2 986 patientes virtuelles synthétiques, remplaçant un essai chez des patientes et démontrant que des évaluations réglementaires peuvent, dans certains cas, s'appuyer sur des cohortes entièrement virtuelles⁴⁶. Les résultats de cette étude étaient cohérents avec ceux d'un essai clinique humain comparable, démontrant ainsi que des essais d'imagerie *in silico* peuvent, dans certains cas, constituer une source fiable de preuves pour l'évaluation réglementaire des dispositifs d'imagerie.

Un autre exemple est celui de l'Université de Virginie, qui a testé un modèle de pancréas artificiel, par le biais d'algorithmes simulant la fonction du pancréas et de différentes stratégies d'insulinothérapie en boucle fermée. Ce modèle a été testé sur une population de 300 patients virtuels atteints de diabète de type 1⁴⁷. En janvier 2008, la FDA a accepté ce simulateur comme substitut aux essais animaux pour le test préclinique de stratégies d'insulinothérapie, ce qui permet de réduire le recours d'études précliniques animales, et d'accélérer le développement du traitement expérimental.

c. Jumeaux numériques

Les jumeaux numériques désignent des représentations virtuelles fidèles et évolutives d'un système biologique donné, qui peut être un patient, un organe, ou même une population – dans ce cas, on parlera de bras in silico. Ces représentations virtuelles sont alimentées en continu par des données réelles et couplées à des modèles qui décrivent le fonctionnement du vivant.

Le jumeau numérique d'un patient combine des données issues des dossiers médicaux, de l'imagerie, de la génomique, des capteurs et des examens fonctionnels avec des modèles biophysiques et des modèles statistiques ou d'apprentissage automatique, afin d'obtenir une réplique virtuelle d'un patient réel. Calibré pour être la réplique virtuelle d'un individu donné, il reflète l'état actuel et l'évolution probable d'une maladie, et peut être "exposé" virtuellement à différentes options thérapeutiques pour simuler bénéfices cliniques et risques.

L'objectif est donc que chaque patient inclus ait sa réplique virtuelle, ce qui permet d'obtenir une comparaison précise, fidèle et évolutive des options thérapeutiques administrées chez un patient et son jumeau virtuel.

Par ailleurs, pendant un essai, un jumeau mis à jour en temps réel peut aider à prédire une dégradation imminente, déclencher des adaptations du protocole ou guider un monitoring plus intelligent, tandis que des "biomarqueurs numériques" dérivés du jumeau quantifient l'état de la maladie avec une précision accrue.

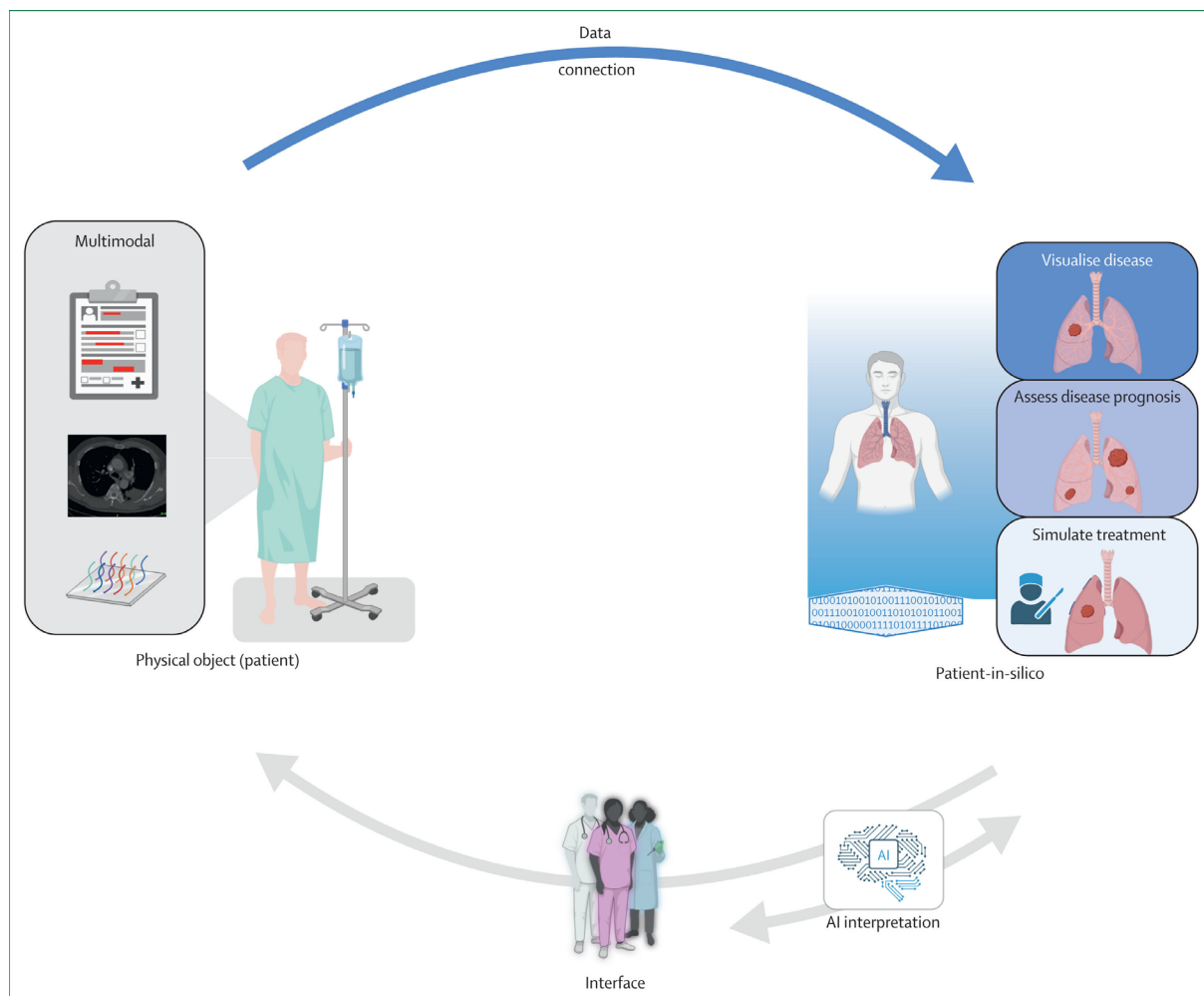


Figure 13. Les jumeaux numériques médicaux⁴⁸

Les avantages potentiels de leur utilisation sont une accélération des étapes de recherche, une diminution des coûts et de la charge pour les patients, une meilleure personnalisation et une puissance statistique accrue grâce à l'intégration massive de données.

Mais cette promesse repose une nouvelle fois sur des exigences fortes : interopérabilité et qualité des données, protection de la vie privée et gouvernance robuste, et surtout crédibilité scientifique des modèles.

Cela implique des processus rigoureux de vérification et de validation, la quantification des incertitudes, des comparaisons systématiques aux données réelles, et une transparence sur les hypothèses et limites.

En neuroradiologie, les jumeaux numériques ont été utilisé pour choisir le meilleur divertisseur de flux - méthode de traitement des anévrismes intracrâniens -

pour un patient donné, avant de mener l'intervention. En effet, le traitement optimal des anévrismes intracrâniens avec des dispositifs de diversion de flux est un enjeu fondamental pour les neuroradiologues ainsi que pour les neurochirurgiens. En raison des formes et des emplacements très irréguliers des anévrismes, le choix du stent et la stratégie de déploiement spécifique au patient peuvent être des décisions très difficiles⁴⁹. L'utilisation des jumeaux numériques a ici permis un meilleur choix du dispositif et une diminution des réinterventions.

Cet exemple montre que les jumeaux numériques ne sont plus seulement une promesse : ils ont déjà servi à concevoir des essais, à remplacer des bras de contrôle, à planifier des interventions et à appuyer des décisions, sous réserve d'une validation rigoureuse, d'une maîtrise des incertitudes et d'une gouvernance adaptée à la protection des données.

V. Conclusion

Nous avons vu au cours de cet ouvrage le potentiel immense de l'utilisation du Big Data dans le secteur de la recherche clinique. De la recherche de candidats-médicaments, à la comparaison des résultats, en passant par la conception de protocoles et l'identification de patients, utiliser les données de santé massives qui sont à notre disposition peut permettre de réduire considérablement le temps, les ressources humaines et financières, mais aussi l'impact des essais cliniques sur les patients. Ce sont autant d'avancées qui pourraient permettre de révolutionner la manière dont les médicaments sont développés par les entreprises pharmaceutiques.

Surtout, ces essais cliniques innovants utilisant les données peuvent être une solution aux limites des essais classiques dans l'ère actuelle de la médecine de précision et du custombuster.

Mais, bien que ces nouvelles technologies semblent pouvoir apporter un bénéfice immense en recherche clinique, la priorité des autorités de santé reste de maintenir la rigueur de la démonstration scientifique et la confiance du public et des professionnels de santé dans le système d'évaluation des médicaments. Il est donc primordial de faire preuve de prudence méthodologique dans ce cadre.

L'utilisation de ces technologies est encore trop peu répandue, au regard de potentiels apports pour les patients et pour la science.

Pour que ces nouvelles technologies puissent être utilisées en pratique dans les essais cliniques, il serait nécessaire que des guidances claires émanent des agences de santé, notamment concernant la méthodologie et le type de bases de données qu'elles attendent, afin de considérer ces essais comme recevables. Sans cela, les industriels ne peuvent pas réellement avancer sur ce terrain. Un cadre doit être co-construit entre les industriels et les autorités de santé pour permettre à ces technologies de voir le jour dans la pratique courante et de se développer.

Plusieurs solutions pourraient permettre d'utiliser et de valider l'utilisation du Big Data dans les essais cliniques, tout en répondant aux exigences requises aujourd'hui par les agences de santé.

Pour commencer, on pourrait envisager une approche visant tester le bras contrôle externe en le comparant au bras contrôle réel afin de démontrer la validité de l'utilisation des cohortes synthétiques dans le contexte étudié. Nous aurions ainsi un essai à 3 bras :

- Un bras expérimental, avec des patients recevant le traitement à évaluer
- Un bras contrôle traditionnel, avec des patients recevant le traitement standard ou un placebo
- Un bras contrôle externe, basé sur des données collectées à partir de bases

Une autre approche à envisager est celle des essais dits émulés. Cette approche consiste à reproduire, à partir de données de vie réelle, les résultats d'un essai clinique déjà conduit. Cette méthode pourrait contribuer à rassurer collectivement sur la validité des approches utilisant des données de vie réelle. Elle permet également aux industriels de se convaincre de la robustesse méthodologique de ces approches, tout en n'ayant pas recours à de nouveaux patients.

Par ailleurs, les autorités de santé semblent vouloir progresser dans le sens de ces nouvelles technologies en santé. Par exemple, la HAS semble aujourd'hui s'orienter vers le recrutement de profils de data scientists et d'ingénieurs en intelligence artificielle, ce qui indique une volonté de développer des compétences internes sur ces sujets.

Bibliographie

1. Allasonniere S, Andre F, Balazard F, Baschet L, Bertoye PH. État d'avancement des évolutions méthodologiques en recherche clinique [Internet]. 2025. Disponible sur: https://www.fcrin.org/sites/default/files/brique/fichier/05-2025/Rapport%20AIS%20GT%20F-CRIN-AIS_Rapport_20250515_vf_0.pdf
2. Quelles sont les différentes phases de développement d'un médicament ? — Notre Recherche Clinique [Internet]. [cité 29 août 2026]. Disponible sur: <https://notre-recherche-clinique.fr/comprendre/#182>
3. Synergie Lyon Cancer. Le parcours du médicament : recherche et développement des médicaments contre le cancer [Internet]. Lyon: Synergie Lyon Cancer; [date de publication non précisée]. Disponible sur: <https://www.synergielyoncancer.fr/les-enjeux/le-parcours-du-medicament>
4. ANSM [Internet]. [cité 29 août 2026]. Bienvenue dans notre rapport d'activité 2024 ! Disponible sur: <https://ansm.sante.fr/qui-sommes-nous/rapport-dactivite-2024/bienvenue-dans-notre-rapport-dactivite-2024>
5. Kassel R. Big Data : définition, technologies, utilisations, formations. Liora [Internet]. 18 févr 2026 [cité 29 août 2026]. Disponible sur: <https://liora.io/big-data-tout-savoir>
6. Austin C, Kusumoto F. The application of Big Data in medicine: current implications and future directions. J Interv Card Electrophysiol. 2016;47(1):51-59. doi:[10.1007/s10840-016-0104-y](https://doi.org/10.1007/s10840-016-0104-y)
7. Big Data 2025 : Les Tendances Incontournables à Surveiller [Internet]. [cité 29 août 2026]. Disponible sur: <https://techreview.fr/actualites-et-tendances-tech/les-tendances-big-data-a-suivre-en-2025/>
8. Haute Autorité de Santé [Internet]. [cité 31 août 2026]. Le service médical rendu (SMR) et l'amélioration du service médical rendu (ASMR). Disponible sur: https://www.has-sante.fr/jcms/r_1506267/fr/le-service-medical-rendu-smr-et-l-amelioration-du-service-medical-rendu-asmr

9. Tableau élaboré par l'autrice, inspiré de : [Haute Autorité de Santé - Le service médical rendu \(SMR\) et l'amélioration du service médical rendu \(ASMR\)](#)
10. Haute Autorité de santé. La HAS et l'accès au remboursement des médicaments [présentation]. 28 mars 2019. Diapositive 29. Disponible sur : <https://fr.slideshare.net/slideshow/le-rle-de-la-has-dans-lacces-au-remboursement-des-medicaments/138619098>
11. 100 questions | Leem [Internet]. [cité 29 août 2026]. Disponible sur: <https://www.leem.org/100-questions>
12. Calameo.com [Internet]. [cité 31 août 2026]. 55. Accès au marché d'un médicament en France : de l'autorisation de mise sur le marché à la décision de remboursement et de prix. Disponible sur: <https://www.calameo.com/read/0020492846434c2bd2ba5>
13. Info E rédactionnelle de N. Comprendre les études scientifiques : les critères qualité incontournables. Nutrixeal Info [Internet]. 1 oct 2020 [cité 29 août 2026]. Disponible sur: <https://nutrixeal-info.fr/comprendre-les-etudes-scientifiques-les-criteres-qualite-incontournables/>
14. Abecassis, P. et Coutinet, N. (2018). V. Du modèle blockbuster au modèle custombuster. Économie du médicament (p. 95-114). La Découverte. <https://shs.cairn.info/economie-du-medicament--9782707196149-page-95?lang=fr>
15. Debievre D. Épidémiologie du cancer du poumon : évolution des vingt dernières années. Bulletin du Cancer. mars 2025;112(3):3S3-15. doi:[10.1016/S0007-4551\(25\)00152-3](https://doi.org/10.1016/S0007-4551(25)00152-3)
16. Fervers B, Clin B, Pairon JC. Épidémiologie du cancer du poumon en France. Revue des Maladies Respiratoires Actualités. oct 2023;15(2):2S5-9. doi:[10.1016/S1877-1203\(23\)00122-2](https://doi.org/10.1016/S1877-1203(23)00122-2)
17. Shi J, Chen Y, Peng C, Kuang L, Zhang Z, Li Y, et al. Advances in Targeted Therapy Against Driver Mutations and Epigenetic Alterations in Non-Small Cell Lung Cancer. Oncologie. 2022;24(4):613-48. doi:[10.32604/oncologie.2022.027545](https://doi.org/10.32604/oncologie.2022.027545)

18. Chu QS. Targeting non-small cell lung cancer: driver mutation beyond epidermal growth factor mutation and anaplastic lymphoma kinase fusion. *Ther Adv Med Oncol.* janv 2020;12:1758835919895756. doi:[10.1177/1758835919895756](https://doi.org/10.1177/1758835919895756)
19. Powley IR, Patel M, Miles G, Pringle H, Howells L, Thomas A, et al. Patient-derived explants (PDEs) as a powerful preclinical platform for anti-cancer drug and biomarker discovery. *Br J Cancer.* 17 mars 2020;122(6):735-44. doi:[10.1038/s41416-019-0672-6](https://doi.org/10.1038/s41416-019-0672-6)
20. Cancer IND. Plateformes de génétique moléculaire : missions et localisation [Internet]. 2022 [cité 29 août 2026]. Disponible sur: <https://en-www.cancer.fr/professionnels-de-sante/parcours-de-soins-des-patients/traitements/traitements-medicamenteux/les-therapies-ciblees/les-plateformes-de-genetique-moleculaire-des-cancers/missions-et-localisation-des-plateformes>
21. Marchetti A, Ferro B, Pasciuto MP, Zampacorta C, Buttitta F, D'Angelo E. NTRK gene fusions in solid tumors: agnostic relevance, prevalence and diagnostic strategies. *Pathologica.* juin 2022;114(3):199-216. doi:[10.32074/1591-951X-787](https://doi.org/10.32074/1591-951X-787)
22. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer.* janv 2009;45(2):228-47. doi:[10.1016/j.ejca.2008.10.026](https://doi.org/10.1016/j.ejca.2008.10.026)
23. Dummerre-Faucoeur A. Thérapie agnostique de tumeurs : exemple du développement d'un traitement ciblant les fusions NTRK [Internet]. Université de Lille (2022-...); 2022 [cité 29 août 2026]. Disponible sur: <https://pepite.univ-lille.fr/ori-oai-search/notice/view/univ-lille-36587>
24. Hong DS, Bauer TM, Lee JJ, Dowlati A, Brose MS, Farago AF, et al. Larotrectinib in adult patients with solid tumours: a multi-centre, open-label, phase I dose-escalation study. *Annals of Oncology.* févr 2019;30(2):325-31. doi:[10.1093/annonc/mdy539](https://doi.org/10.1093/annonc/mdy539)

25. Bayer. A Study to Learn How Well the Drug Larotrectinib Works in Adults With Different Solid Cancers With a Change in the Genes Called NTRK Fusion [Clinical trial registration] [Internet]. clinicaltrials.gov; juin 2026 [cité 29 août 2026]. Clinical trial registration no.: NCT02576431. Disponible sur: <https://clinicaltrials.gov/study/NCT02576431>
26. Institut national du cancer. L'accès aux nouveaux traitements [Internet]. Boulogne-Billancourt: Institut national du cancer; [date de publication non précisée; mis à jour le 17 juin 2022]. Disponible sur: <https://www.cancer.fr/personnes-malades/parcours-de-soins/principaux-traitements/medecine-de-precision/acces-aux-nouveaux-traitements>
27. Laetsch TW, DuBois SG, Mascarenhas L, Turpin B, Federman N, Albert CM, et al. Larotrectinib for paediatric solid tumours harbouring NTRK gene fusions: phase 1 results from a multicentre, open-label, phase 1/2 study. The Lancet Oncology. mai 2018;19(5):705-14. doi:[10.1016/S1473-2045\(18\)30119-0](https://doi.org/10.1016/S1473-2045(18)30119-0)
28. Haute Autorité de Santé [Internet]. [cité 31 août 2026]. VITRAKVI (larotrectinib). Disponible sur: https://www.has-sante.fr/jcms/p_3193768/fr/vitrakvi-larotrectinib
29. ISO [Internet]. [cité 29 août 2026]. Les dossiers électroniques du patient en bref. Disponible sur: <https://www.iso.org/fr/soins-sante/dossier-electronique-patient>
30. Velcheti V, Hu X, Yang L, Pietanza MC, Burke T. Long-Term Real-World Outcomes of First-Line Pembrolizumab Monotherapy for Metastatic Non-Small Cell Lung Cancer With ≥50% Expression of Programmed Cell Death-Ligand 1. Front Oncol. 25 mars 2022;12:834761. doi:[10.3389/fonc.2022.834761](https://doi.org/10.3389/fonc.2022.834761)
31. Le système national des données de santé (SNDS) et l'accès aux données de santé | Direction de la recherche, des études, de l'évaluation et des statistiques [Internet]. [cité 29 août 2026]. Disponible sur: <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/le-systeme-national-des-donnees-de-sante-snds-et-lacces-aux-donnees-de>

32. 1000 Genomes | A Deep Catalog of Human Genetic Variation [Internet]. [cité 31 août 2026]. Disponible sur: <https://www.internationalgenome.org>
33. pharmaphorum [Internet]. 2021 [cité 31 août 2026]. BMS embraces AI drug discovery with \$1.2bn Exscientia alliance. Disponible sur: <https://pharmaphorum.com/news/bms-embraces-ai-drug-discovery-with-1-2bn-exscientia-alliance>
34. businesswire.com/news/home/20260717090949/en/Aqemia-and-Sanofi-Expand-Their-Research-Collaboration [Internet]. [cité 31 août 2026]. Disponible sur: <https://www.businesswire.com/news/home/20260717090949/en/Aqemia-and-Sanofi-Expand-Their-Research-Collaboration>
35. Burki T. A new paradigm for drug development. The Lancet Digital Health. mai 2020;2(5):e226-7. doi:[10.1016/S2589-7500\(20\)30088-1](https://doi.org/10.1016/S2589-7500(20)30088-1)
36. Artificial intelligence yields new antibiotic | MIT News | Massachusetts Institute of Technology [Internet]. [cité 31 août 2026]. Disponible sur: <https://news.mit.edu/2020/artificial-intelligence-identifies-new-antibiotic-0220>
37. Exscientia plc. Centaur Chemist™: Design Make Test [illustration]. Oxford (UK): Exscientia; [date inconnue]. Disponible sur: <https://www.exscientia.ai/>.
38. Pioneering AI Drug Discovery | Recursion [Internet]. [cité 31 août 2026]. Disponible sur: <https://www.recursion.com>
39. RECOVERY Trial [Internet]. [cité 31 août 2026]. Welcome. Disponible sur: <https://www.recoverytrial.net>
40. Duke University. Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-term Effectiveness [Clinical trial registration] [Internet]. clinicaltrials.gov; juin 2021 [cité 31 août 2026]. Clinical trial registration no.: NCT02697916. Disponible sur: <https://clinicaltrials.gov/study/NCT02697916>

41. Wierzbicki J. Essais cliniques - Comparaison virtuelle [Internet]. Disponible sur: <https://www.unicancer.fr/wp-content/uploads/sites/2/2021/10/phq295-prospectiveenquete-brasvirtuelsjw.pdf>
42. Burger HU, Gerlinger C, Harbron C, Koch A, Posch M, Rochon J, et al. The use of external controls: To what extent can it currently be recommended? *Pharmaceutical Statistics*. nov 2021;20(6):1002-16. doi:[10.1002/pst.2120](https://doi.org/10.1002/pst.2120)
43. Wang S, Kidwell KM, Roychoudhury S. Dynamic Enrichment of Bayesian Small-Sample, Sequential, Multiple Assignment Randomized Trial Design Using Natural History Data: A Case Study from Duchenne Muscular Dystrophy. *Biometrics*. 1 déc 2023;79(4):3612-23. doi:[10.1111/biom.13887](https://doi.org/10.1111/biom.13887)
44. Etude Essais cliniques 2030 - La révolution de la recherche clinique [Internet]. [cité 29 août 2026]. Disponible sur: <https://www.leem.org/publication/etude-essais-cliniques-2030-la-revolution-de-la-recherche-clinique>
45. Druart L, Graham Longworth S, Rolland C, Dolgoploff M, Terrisse H, Bosson JL, et al. Can an Open-Label Placebo Be as Effective as a Deceptive Placebo? Methodological Considerations of a Study Protocol. *Medicines*. 2 janv 2020;7(1):3. doi:[10.3390/medicines7010003](https://doi.org/10.3390/medicines7010003)
46. Badano A, Graff CG, Badal A, Sharma D, Zeng R, Samuelson FW, et al. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. *JAMA Netw Open*. 30 nov 2018;1(7):e185474. doi:[10.1001/jamanetworkopen.2018.5474](https://doi.org/10.1001/jamanetworkopen.2018.5474)
47. Cobelli C, Kovatchev B. Developing the UVA/Padova Type 1 Diabetes Simulator: Modeling, Validation, Refinements, and Utility. *J Diabetes Sci Technol*. nov 2023;17(6):1493-505. doi:[10.1177/19322968231195081](https://doi.org/10.1177/19322968231195081)
48. MEDVASC - Jumeaux numériques médicaux [Internet]. [cité 31 août 2026]. Disponible sur: <https://medvasc.info/archives-blog/jumeaux-num%C3%A9riques-m%C3%A9dicaux>

49. Janiga G, Daróczy L, Berg P, Thévenin D, Skalej M, Beuing O. An automatic CFD-based flow diverter optimization principle for patient-specific intracranial aneurysms. *Journal of Biomechanics*. nov 2015;48(14):3846-52. doi:[10.1016/j.jbiomech.2015.09.039](https://doi.org/10.1016/j.jbiomech.2015.09.039)

Université de Lille
UFR3S-Pharmacie
DIPLOME D'ETAT DE DOCTEUR EN PHARMACIE
Année Universitaire 2026/2027

Nom : PIETTRE
Prénom : Valentine

Titre de la thèse : Utilisation du big data en recherche clinique : Comment les données massives peuvent-elles optimiser la conception, la conduite et les résultats des essais cliniques ?

Mots-clés : recherche clinique, médecine de précision, évolution, thérapie ciblée, agences réglementaires, accès au marché, innovation, base de données, apprentissage automatique, intelligence artificielle, éthique, protection des données

Résumé : Le système actuel d'évaluation des médicaments est confronté à plusieurs défis majeurs, tels que l'essor de la médecine de précision, les enjeux éthiques et l'ambition de réduire toujours davantage le temps et les coûts liés à la recherche et au développement. Le Big Data offre des solutions pour répondre à ces problématiques, via des approches novatrices, telles que les cohortes synthétiques, les bras virtuels et les jumeaux numériques, à partir de données collectées via des sources diverses. Il est cependant important de rappeler que ces approches novatrices doivent permettre de réduire les contraintes citées précédemment, tout en maintenant des normes élevées de sécurité, d'efficacité et de démonstration des effets, afin de satisfaire aux exigences de autorités d'évaluation des médicaments, et de garantir l'accès à des traitements fiables et de qualité pour les patients.

Membres du jury :

Président : Monsieur Mohamed Lemdani, Professeur des Universités en Biomathématiques, Université de Lille

Directeur de thèse : Professeur Pascal DAO PHAN, Directeur de la Recherche clinique France/Benelux chez Bayer HealthCare SAS à Lille, Professeur Associé de à la Faculté de Pharmacie (UFR3S, Université de Lille).

Membre(s) extérieur(s) : Monsieur Jean-Baptiste Colin, Responsable médical oncologie, Bayer Lille