

État d'avancement des évolutions méthodologiques en recherche clinique

De la recherche expérimentale au suivi d'évaluation en phase d'usage courant

GT AIS/F-CRIN « *L'évolution des méthodologies d'essais cliniques : nouveaux outils, nouveaux usages et conditions de recours* »

Rapport
Mai 2025

Ce rapport est le fruit d'un travail collectif qui a mobilisé pendant deux ans un groupe de 31 experts sur les « *évolutions méthodologiques en recherche clinique* » mis en place par l'Agence de l'innovation en santé (AIS) et l'infrastructure F-CRIN (French Clinical Research Infrastructure Network).

Des remerciements chaleureux leur sont adressés pour leur participation active, leur contribution et leur mobilisation sans faille malgré leurs contraintes professionnelles.

Stéphanie ALLASSONNIERE	Professeur de mathématiques à l'UFR de médecine, Chaire de recherche et Directrice adjointe de l'institut PR[AI]RIE, Vice-Présidente valorisation d'UPCité
Fabrice ANDRE	Directeur de la recherche – Institut Gustave Roussy
Félix BALAZARD	Director of Optimized Development – Société OWKIN
Louise BASCHET	Co-fondatrice – HORIANA
Pierre-Henri BERTOYE	UNICANCER, directeur de la prospective réglementaire et DGS, chargé de mission sur des dossiers européens recherche et innovation
Olivier BLIN	AMU-AP-HM – délégation Tafalgie Therapeutics
Michel CUCHERAT	CEO – Société MetaEvidence ; co investigateur
Marc CUGGIA	Professeur en informatique médicale – Coordonnateur du LabCom LITIS – Université Rennes 1
Manon DE FALLOIS	Adjointe à la cheffe du service santé – CNIL
Olivier DEMARCQ	Directeur Médical Maladies Rares – Pfizer International (au titre du Leem)
Dominique DEPLANQUE	Directeur du CIC 1403 Inserm – CHU et Université de Lille – Coordonnateur du réseau national des CIC Inserm
Hélène ESPEROU	Responsable du pôle de Recherche clinique – Inserm
Lionel DA CRUZ	Bureau Organisation et financement de la recherche (RI1) - Ministère de la Santé et de la Prévention
Jean-Louis FRAYSSE	Directeur associé – BOTdesign
Lamia GRIMALDI	Professeur d'épidémiologie et santé publique – AP-HP (Hôpital Ambroise Paré)
Stéphanie KERVESTIN	Déléguée générale – ARIIS
Gil LAMPE	Directeur des données – Health Data Hub
Pierre LEMAITRE	Clinical trials assessor – ANSM
Edouard LHOMME	Plateforme EUCLID – Université de Bordeaux
Benjamin MATEI DEDIU	Agence de l'Innovation en Santé
Emmanuel PHAM	Société Nova in Silico
David PEROL	Directeur de la recherche clinique et de l'innovation – CLCC Léon Bérard
Raphaël PORCHER	AP-HP – Université Paris Cité
Werner REIN	Directeur médical, Société Theranexus
Mathieu ROBAIN	Directeur scientifique – Direction des DATA – Unicancer
Matthieu ROUSTIT	Professeur associé - Université Grenoble Alpes
Bénédicte SIMON	Directrice de la recherche et de l'innovation – CHU de Brest
Hélène VIATGE	CEO – Société Agora Health
Stéphane VIGNOT	CLCC Reims
Philippe ZAMIA	Référent méthodologie – Direction Europe & innovation – ANSM
Sarah ZOHAR	Directrice de l'U1346 'HeKA – INRIA, Inserm, UPCité

Des remerciements sont également adressés aux contributeurs et relecteurs de ce document dont les observations, leurs suggestions ont permis d'enrichir le document dont :

Lise Alter, Clémentine Body, Rémy Choquet, Charles-Edouard Escurat, Jérôme Fabiano, Line Farah, Thomas Filleron, Laetitia Gambotti, Vincent Gazin, Benoit Labarthe, Franck Mouthon, Christine Trillou, Viet-Thi Tran, ainsi que les adhérents d'ARIIS, de la FIAC et du Leem

Les copilotes du GT AIS /F-CRIN :

- Camille SCHURTZ, Chargée de mission à l'Agence de l'Innovation en Santé
- Vincent DIEBOLT, Directeur de l'infrastructure « F-CRIN »

Résumé

La recherche clinique est au cœur de l'évaluation de la sécurité et de l'efficacité des produits ou de technologies de santé. Elle marque le passage d'une découverte scientifique à une innovation validée pour un usage médical, en mesurant le bénéfice/risque et les conditions d'utilisation.

Ce rapport, fruit d'une réflexion lancée en avril 2023 par l'Agence de l'innovation en santé (AIS) et l'infrastructure *French Clinical Research Infrastructure Network* (F-CRIN), vise à explorer l'impact des évolutions méthodologiques en recherche clinique par rapport au modèle de référence : l'essai contrôlé randomisé. Il s'appuie sur les contributions d'experts issus de divers domaines (médecine, statistique, droit, informatique), enrichies de travaux conduits lors d'auditions, d'ateliers de réflexion (comme les Ateliers de Giens) ou dans le cadre d'associations comme SILICA.

Les travaux ont permis 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, ceci s'intégrant dans le contexte global du foisonnement de l'innovation (ou présentée comme telle) diagnostique et thérapeutique.

Trois grands facteurs justifient ces évolutions :

- La médecine des 5P (Personnalisée, Préventive, Prédicative, Participative et fondée sur les Preuves) impose une évaluation plus fine de traitements ciblés pour des sous-populations, compliquant le recrutement.
- Le développement de bases de données massives facilite l'analyse en conditions réelles d'utilisation.
- L'intégration de l'intelligence artificielle (IA) et de données dites artificielles permet d'enrichir les analyses et d'accélérer l'évaluation.

Le rapport s'articule autour des axes suivants :

- L'utilisation des données de soins courant pour soutenir l'évaluation clinique (section 0, présentant l'approche d'émulation d'essais cliniques, et discutant des cas d'usage des essais avec bras contrôle externe et essais augmentés).
- L'apport de l'IA et des données artificielles pour créer des modèles pronostiques ou simuler des essais (section 4, présentant les applications de l'intelligence artificielle à l'optimisation des essais cliniques, les modèles pronostics, et l'utilisation de la simulation et du *in silico* dans la recherche clinique).

Les exemples présentés montrent à la fois l'intérêt croissant pour ces nouvelles approches et les exigences élevées pour leur validation.

L'évaluation des évolutions méthodologiques a été guidée par l'impératif d'assurer un niveau de preuve comparable à l'approche basée sur l'essai contrôlé randomisé, et d'apporter aux patients le même niveau de sécurité. Les évolutions méthodologiques n'ont pas été appréhendées comme des alternatives au modèle en place mais plutôt sous l'angle de modèles complémentaires apportant des éléments de preuve additionnels, afin de renforcer la puissance statistique ou de pallier les limites et les obstacles auxquels le modèle de référence peut faire face.

Des cas pratiques axés sur la démonstration de la valeur de ces outils sont indispensables pour avancer sur l'acceptabilité de ces méthodes et outils par les régulateurs.

Enfin, les interactions avec les régulateurs sont cruciales. Le rapport s'inscrit dans une dynamique de dialogue avec l'Agence National de Sécurité du Médicament (ANSM), la Haute Autorité de la Santé (HAS) et les initiatives européennes, comme l'*Accelerating Clinical Trials in the European Union* (ACT-EU). Des dispositifs d'accompagnement existent, comme le Guichet Innovation à l'ANSM ou les entretiens HAS pour le forfait innovation, et les rencontres précoces HAS pour les spécialités pharmaceutiques. Côté européen, l'*European Medicines Agency* (EMA) a mis en place une procédure de qualification des méthodes

innovantes, tandis que la Food and Drug Administration (FDA) soutient également ces démarches via le CDER Center for Clinical Trial Innovation (C3TI).

Table des matières

Résumé _____	4
1. Analyse du contexte et vocation du rapport _____	7
2. Méthode _____	8
3. Comment utiliser les données de soins courants pour étayer les décisions ? _____	9
3.1. Intérêt croissant des données dans l'évaluation initiale, et dans le suivi de l'efficacité en vie réelle.	9
3.2. Cadre conceptuel : l'émulation d'essai clinique _____	10
3.3. Pour une évaluation initiale : designs et cas d'usage _____	13
3.3.1. Essais avec bras de contrôle externe _____	13
3.3.2. Essais augmentés par données collectées en pratique courante _____	16
3.4. Pour le suivi de l'efficacité en vie réelle _____	18
3.5. Recommandations concernant les sources de données collectées en pratique courante _____	19
4. Le numérique et l'IA au service de la recherche clinique _____	21
4.1. Intérêt des outils numériques pour accélérer et optimiser la conception, la réalisation et l'analyse des essais cliniques. _____	21
4.2. Designs et cas d'usage _____	21
Applications de l'intelligence artificielle à l'optimisation des essais cliniques _____	21
Modèles pronostics _____	23
Utilisation de la simulation et du <i>in silico</i> dans la recherche clinique _____	25
5. Conclusion _____	30
6. Bibliographie _____	30

1. Analyse du contexte et vocation du rapport

La recherche clinique est au cœur de l'évaluation de la sécurité et de l'efficacité des produits de santé. Elle marque le passage d'une découverte scientifique à une innovation validée pour un usage médical, en mesurant le bénéfice/risque et les conditions d'utilisation.

Dans la vocation du rapport Innovation Santé 2030 et des engagements du Conseil Stratégique des Industries de Santé (CSIS), il est essentiel de faciliter et d'accélérer la recherche biomédicale, tout en garantissant le plus haut niveau de sécurité pour les participants aux études et la fiabilité des résultats produits.

Les principaux enjeux résident dans la nécessité de dépasser les limites et les contraintes connues du schéma traditionnel de l'essai contrôlé randomisé par étapes successives, notamment en termes de difficultés de mobilisation et de recrutement de volontaires et de patients dans les délais prévus. Bien que des méthodes complémentaires et/ ou alternatives existent, elles ne sont pas encore largement reconnues ni validées par les agences de régulation et restent peu appropriées par les porteurs d'innovation.

De plus, la mission « Régulation des produits de santé », consistant à identifier les voies d'évolution possibles du système de financement et régulation des produits de santé, a reconnu pleinement la place de l'évaluation en vie réelle dans le parcours de développement, de mise à disposition, et d'évaluation continue de l'innovation (1).

L'enjeu est donc de démontrer la performance et la valeur ajoutée de ces nouvelles approches méthodologiques à des moments clés de l'évaluation clinique des produits de santé, par une évaluation comparative et le suivi de cas d'étude.

Dans cette perspective, des travaux communs ont été engagés entre l'Agence de l'Innovation en Santé (AIS) et l'infrastructure *French Clinical Research Infrastructure Network* (F-CRIN), notamment via la constitution d'un groupe de travail (GT), afin d'apporter des réponses concrètes à ces enjeux méthodologiques.

L'objectif principal du GT AIS/F-CRIN est d'évaluer la pertinence de ces nouvelles méthodologies en les comparant aux méthodes classiques, puis de définir les cas dans lesquels il est possible d'y avoir recours et d'identifier les éventuels freins à lever. Plus précisément, il s'agit d'étendre le champ de réflexion à l'ensemble des évolutions méthodologiques et opérationnelles (designs modernes, utilisation des données de vie réelle, essais *in silico*, essais décentralisés), en lien avec les travaux européens (*European Medicines Agency* (EMA), *European Network for Health Technology Assessment* (EUnetHTA)) et internationaux (*Food and Drug Administration* (FDA)). Le GT ambitionne de contribuer à la reconnaissance effective de ces nouvelles approches par les agences de régulation notamment par la production de rapports sur ces dernières.

Ce rapport complète les deux précédents rapports axés sur l'état des lieux des méthodologies innovantes (2) ainsi que le rapport axé « Pour l'utilisation des données de santé en vie réelle en recherche clinique » (3). Ce rapport s'articule autour de 2 axes :

1. L'utilisation des données de soins ou collectées dans un cadre épidémiologique ou de recherche pour soutenir l'évaluation clinique, dans le cadre de l'évaluation initiale (ou dans une extension d'indication), et dans le cadre du suivi de l'efficacité en vie réelle. ¹
2. L'apport de l'Intelligence Artificielle (IA) et des données synthétiques pour créer des modèles pronostiques ou simuler des essais.

¹ *Nota bene* : Ce document n'a pas pour vocation de reprendre l'ensemble des éléments du guide méthodologique des études en vie réelle de la Haute Autorité de la Santé (HAS) (4), dont le champ d'application couvre d'autres objectifs que l'évaluation de l'efficacité.

Ce rapport s'adresse à un large panel d'acteurs, incluant les professionnels de santé, les porteurs d'innovation, les promoteurs de recherche, les patients ainsi que les structures publiques concernées.

2. Méthode

Le rapport est le résultat de la réflexion conduite par un groupe de travail mis en place en avril 2023 à l'initiative de l'AIS et de F-CRIN. Ce collectif comprenait 31 experts aux compétences complémentaires (méthodologiques, statistiques, réglementaires) et d'appartenance professionnelle très divers (hospitalier, académiques, industriels en santé et du numérique), assurant une représentativité de l'ensemble des acteurs et parties prenantes et une garantie de livrables transposables en pratique sur les méthodologies d'essais cliniques.

Après une 1^{ère} étape d'état des lieux et d'analyse des méthodologies alternatives de recherche clinique, le collectif qui s'est réuni à intervalles réguliers et a procédé à plus de trente auditions, a été scindé en trois sous-groupes thématiques, chacun étant coanimé par un binôme d'experts afin de poursuivre la réflexion au travers de trois angles différents :

1. La gestion de la donnée, qui constitue la brique de soutènement des essais cliniques
2. Le recours aux méthodes alternatives d'essais cliniques en phase initiale d'évaluation pour la mise sur le marché et le remboursement pour accélérer le développement et enrichir le niveau de preuve
3. L'usage de ces méthodes en phase d'aval dans les conditions d'utilisations, précoce et au long cours pour mesurer la performance en vie réelle.

Sur la base de ces réflexions, deux livrables ont d'ores et déjà été produits :

- En mars 2024 une analyse des avis rendus par les organismes régulateurs (FDA, EMA) et l'autorité évaluatrice (HAS) (2)
- Un rapport en mai 2024 sur « *L'utilisation des données de santé en vie réelle en recherche clinique : état des lieux, analyse et recommandations* » (3)

Le présent rapport constitue une 3^{ème} contribution avec un focus sur l'usage des données de soins ou collectées dans un cadre épidémiologique ou de recherche et l'apport de l'IA et des données synthétiques pour créer des modèles pronostiques ou simuler des essais.

Ce rapport pourra être complété ou actualisé.

3. Comment utiliser les données de soins courants pour étayer les décisions ?

3.1. Intérêt croissant des données dans l'évaluation initiale, et dans le suivi de l'efficacité en vie réelle.

Dans le cadre d'une évaluation initiale (ou l'évaluation dans une nouvelle indication), dans certaines situations, l'essai contrôlé randomisé (ECR) peut ne pas être réalisable pour des raisons éthiques ou d'acceptabilité, ou pourrait être complexe à réaliser dans des délais raisonnables (maladies rares et sous-types moléculaires en oncologie). L'impossibilité est justifiée par le promoteur de l'étude.

Il peut également arriver que l'ECR soit bien réalisé mais lacunaire (le suivi à court terme nécessite une confirmation sur le long terme, les analyses des sous-groupes présentent un trop faible nombre de patients, le comparateur n'est plus le comparateur pertinent). En effet, les essais cliniques, étant intrinsèquement limités sur les trois axes que sont le temps, la diversité des patients et l'exposition aux traitements existants, ne permettent parfois pas de lever certaines incertitudes qui peuvent peser sur l'impact des nouveaux traitements à moyen et long-terme, sur l'ensemble de la population cible, par rapport au standard de prise en charge. Ainsi, par exemple, la construction d'un bras comparateur externe peut venir en complément d'un essai contrôlé randomisé classique, afin de permettre à but exploratoire la comparaison avec un autre bras, voire plusieurs bras externes, en fonction de la prise en charge standard recommandée, qui peut être différente dans les pays.

Dans le cadre d'un suivi de l'efficacité en vie réelle, des études observationnelles de « *comparative effectiveness* », l'efficacité en vie réelle, peuvent être utiles. Par leur nature elles ne sont pas adaptées au développement initial d'un nouveau traitement (qui n'est pas encore utilisé en pratique courante). Elles peuvent cependant être envisagées pour lever des incertitudes après l'enregistrement et l'accès au marché.

De plus, les essais pragmatiques ont fait l'objet de discussions par le groupe de travail sur l'utilisation de soin courant pour étayer les décisions concernant un produit ou une technologie de santé. Les essais pragmatiques sont de véritables essais randomisés, mais qui s'appuient sur les structures habituelles de recueil d'information pour récupérer leurs données. Il ne s'agit pas à proprement parler d'une nouvelle méthodologie, mais simplement d'une optimisation de la réalisation des essais randomisés. Ainsi, ces designs sont discutés en annexe.

L'ensemble de ces approches sont résumées par Concato et al.(5) et présentées dans la figure ci-dessous.

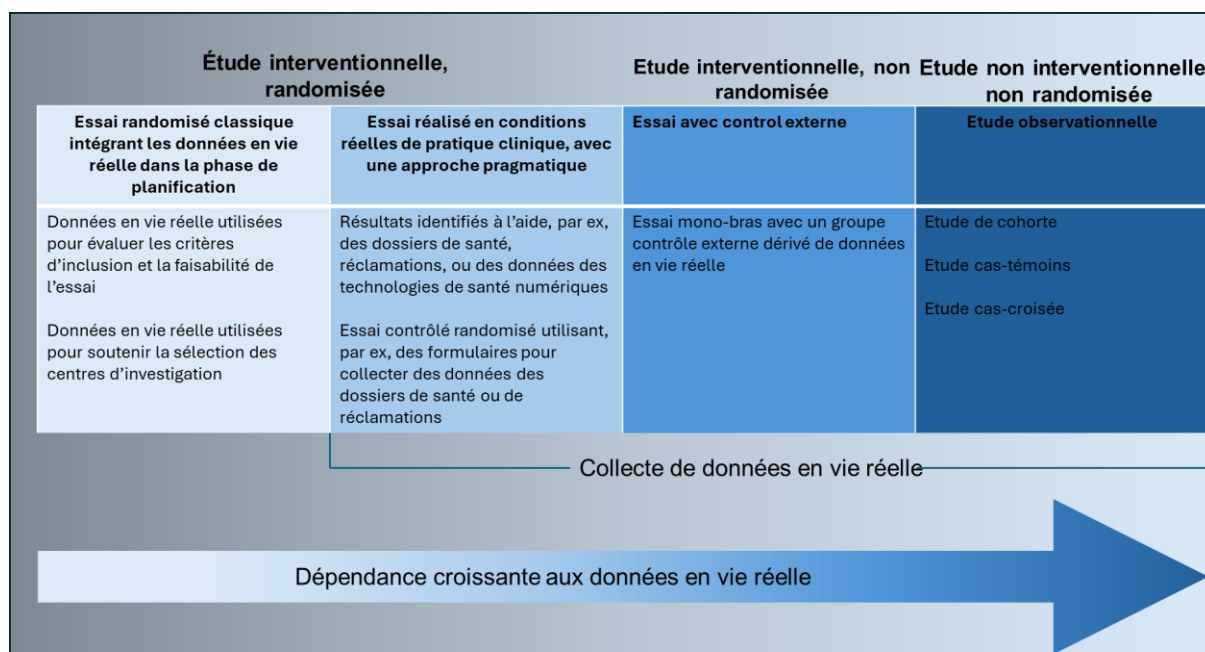


Figure 1 - Dépendance aux données en vie réelle selon la conception de l'étude. Adapté de Concato et al (5)

Dans ce contexte, l'émulation d'essai clinique cible (« *target trial emulation* »)(6), s'impose comme une approche méthodologique rigoureuse permettant de tirer parti des données observationnelles pour estimer un effet causal, dans le prolongement ou en l'absence d'un essai randomisé (présenté section 3.2).

3.2. Cadre conceptuel : l'émulation d'essai clinique

L'émulation d'un essai clinique (6) vise à reproduire, à partir de données observationnelles, ce qui aurait été fait dans un essai randomisé, afin d'estimer un effet causal. Cela implique de définir clairement les critères d'éligibilité, les interventions comparées, la durée de suivi, les critères de jugement et l'allocation au traitement, en utilisant un estimand causal.

La notion d'estimand est un concept statistique et méthodologique qui décrit de manière claire et précise ce que l'on souhaite estimer dans une étude ou un essai clinique : la population d'intérêt, le critère de jugement, l'intervention et le comparateur (si pertinent), la quantification de l'effet d'intérêt, et la stratégie de gestion des événements intercurrents (les événements susceptibles de se produire entre le début de l'étude et l'évaluation).

Hernán et al. (7) ont récemment publié un cadre complet de l'émulation des essais cibles, en faisant la part de ce qui sera de la spécification, et de ce qui sera de l'émulation. Cela permet de décrire précisément les étapes à suivre dans la construction de son protocole et de faire correspondre à chaque étape une cartographie des données utilisées pour émuler. A chaque étape, lorsque la cartographie des données s'éloigne des spécifications du protocole (voir le

Tableau 1, une analyse de sensibilité et une analyse de risque sont requises pour apprécier l'impact de l'écart.

Tableau 1 : Exemple de spécification et d'émulation d'un essai cible, l'émulation étant basée sur des données existantes : Les contrastes causaux d'intérêt sont l'effet de l'affectation et l'effet par protocole, adapté de Hernan et al. 2025 (7)

	Spécification	Émulation
Estimand causal	Critères d'éligibilité Stratégies de traitement Attribution (non masquée) Résultats Début/fin du suivi Contrastes causaux§	Correspondance des données pour chaque critère† Correspondance des données pour chaque composant Correspondance des données pour l'attribution‡ Correspondance des données pour chaque résultat Identique Analogues observationnels des contrastes causaux
Hypothèses d'identification*	Pour l'effet de l'attribution (intention de traiter) : Attribution randomisée Si applicable, hypothèse d'échangeabilité conditionnelle pour : <ul style="list-style-type: none"> • Perte de suivi (lister les facteurs) • Événements concurrents (lister les facteurs) Pour l'effet per protocole : Hypothèse d'échangeabilité conditionnelle (lister les variables confusionnelles de base/temporelles)	Hypothèse d'échangeabilité conditionnelle** Lister les variables de confusion de base et décrire la correspondance des données pour chacune Correspondance des données pour chaque facteur Correspondance des données pour chaque facteur Correspondance des données pour chaque facteur de confusion
Estimateur	Pour chaque estimand causal : analyse de données, y compris analyses de sous-groupes et hypothèses de modélisation	Décrire les modifications requises pour l'émulation, le cas échéant, et les analyses de sensibilité

Les hypothèses nécessaires pour identifier l'effet, même avec une taille d'échantillon infinie. Dans cet exemple, l'échangeabilité conditionnelle est l'hypothèse d'identification pour l'effet de l'attribution.

† Sauf la volonté de participer à une expérience, qui en pratique est un critère d'éligibilité pour l'essai cible mais pas pour son émulation.

‡ Par exemple, prescription ou délivrance du traitement.§ En incorporant les événements concurrents, lorsque c'est approprié. L'estimand causal suppose implicitement qu'il n'y a pas de perte de suivi, donc l'hypothèse est qu'aucune perte de suivi ne correspond à une intervention bien définie.

§ En incorporant les événements concurrents, lorsque c'est approprié. L'estimand causal suppose implicitement qu'il n'y a pas de perte de suivi, donc l'hypothèse est qu'aucune perte de suivi ne correspond à une intervention bien définie.

|| Aucun ajustement nécessaire, mais un ajustement peut être justifié en cas de déséquilibre aléatoire des facteurs de base.

** L'hypothèse selon laquelle les facteurs de confusion de base sont connus et mesurés.

Les différentes étapes sont :

- L'élaboration du protocole de l'essai contrôlé randomisé contrôlé cible : définir les critères d'éligibilité des sujets, stratégies thérapeutiques, procédure d'attribution des traitements, période de suivi, critère de jugement, effet causal, plan d'analyse statistique... De plus, la démarche de gestion des biais dans l'analyse prévue doit être détaillée dans le protocole de l'essai cible.
- L'identification des facteurs de confusion et modificateurs d'effet du traitement : il s'agit d'une étape cruciale. Elle s'opère grâce à une revue systématique de la littérature, la construction de diagrammes de causalité (DAG), et une cotation du niveau de risque de biais.
- L'identification de la ou des bases à utiliser : la sélection objective de la ou des sources de données à utiliser est basée sur les caractéristiques définies a priori dans le protocole. La base de données en vie réelle doit être de forte qualité et granularité (cela sera essentiel pour appliquer les critères d'inclusion et permettre l'analyse des facteurs de confusion), la plus exhaustive possible dans l'indication (pour augmenter la représentativité à la population cible, et la puissance statistique, et disposer des autorisations réglementaires (en particulier être conforme aux prérequis de la Commission Nationale Informatique et Liberté (CNIL)).
- La validation des bases de données : les données doivent être considérées à jour avec un processus a priori de validation des données transparent. Une validation des données présentes dans la base de données avec les dossiers sources est souvent indispensable (cf. dernier guidelines FDA (8)) Les analyses de qualification de la ou des sources de données doivent

impérativement être réalisées en dehors de toutes analyses inférentielles ou indirectement inférentielles (cf. PRINCIPLED (9)).

En particulier, le nombre d'observations doit être suffisant dans le bras de contrôle en vie réelle. En effet, l'effectif de la comparaison doit permettre de calculer une puissance post-hoc suffisante avec les hypothèses utilisées pour bâtir un ECR.

- La réalisation des analyses statistiques rigoureuses (inférence causale par pondération de l'inverse de la probabilité de traitement (*Inverse Probability of Treatment Weighting* (IPTW), g-computation, matching, etc.) afin de limiter les biais de confusion sur les facteurs mesurés. Les biais de sélection et notamment d'immortalité peuvent être limités par design avec la synchronisation du « temps zéro » par émulation d'essai cible, et l'analyse, par l'utilisation de méthodes appropriées (*cloning-censoring-weighting*). L'imputation éventuelle des données manquantes (technique d'imputation multiple) doit parfois être envisagée et les hypothèses sous-jacentes évaluées.
- L'importance du biais de confusion résiduelle nécessite d'être quantifiée (Analyse des Biais Quantitatifs, E-Value, contrôles négatifs pouvant nécessiter un chaînage à des données externes, par exemple aux données du Système National de Données de Santé (SNDS)) (10,11).

Un avantage clé de l'émulation d'un essai clinique est de fixer le début du suivi lorsque les participants remplissent les critères d'éligibilité et commencent le traitement, permettant de synchroniser les groupes et réduire des biais de sélection, comme le temps d'immortalité.

Sa robustesse a été documentée dans des publications visant à titre de preuve de concept à répliquer la taille d'effet obtenue dans des essais cliniques déjà réalisés (12,13). Son intérêt est décrit notamment dans le Livre Blanc de la Société Française de Pharmacologie Thérapeutique (SFPT), «De la nécessité de la méthodologie dans l'évaluation des médicaments» (14), et plus récemment dans les récentes conclusions d'un des ateliers de Giens 2024 (15).

3.3. Pour une évaluation initiale : designs et cas d'usage

Ces études peuvent être de différents types, avec différents niveaux de preuve et leviers encore existants. Sont présentés dans ce document :

- Les études interventionnelles non randomisées, dites monobras, intégrant un contrôle externe.
- Les essais randomisés hybrides (externally-augmented trials), dont le bras contrôle randomisé est complété par des données collectées en pratique courante.

3.3.1. Essais avec bras de contrôle externe

Définition

Ces études intègrent un bras de contrôle externe construit à partir de données collectées en pratique courante, intégrant ou non une mise à jour prospective des données, et utilisent des méthodes appropriées afin d'appréhender les biais, notamment par l'émulation de tout ou partie d'un essai clinique combinée à des méthodes d'ajustement statistique appropriées.

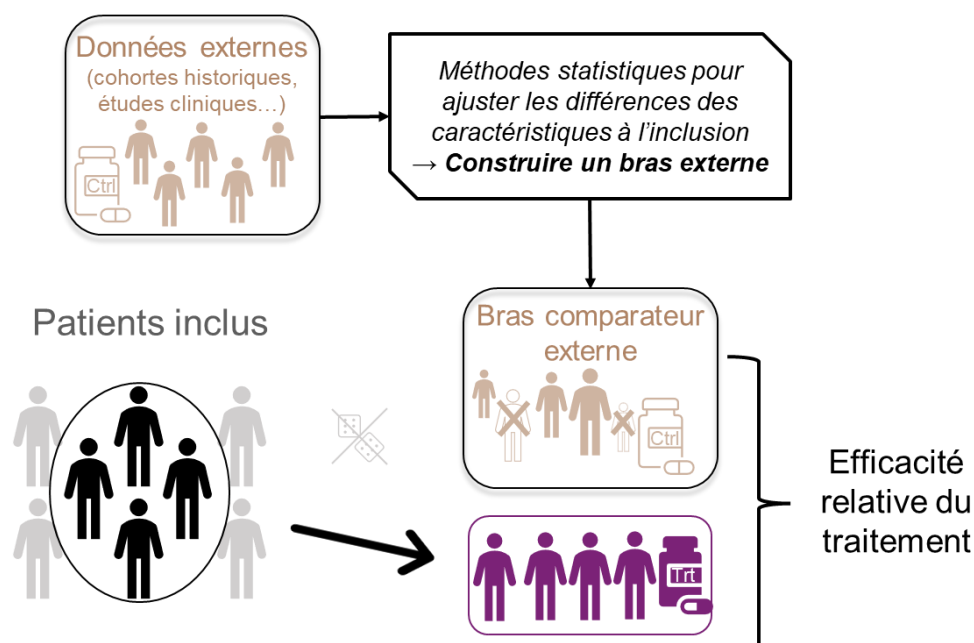


Figure 2 – Illustration d'une étude non randomisée avec bras comparateur externe.

Intérêt

En suivant une méthodologie rigoureuse, les essais avec bras de contrôle externe présentent, entre autres, les intérêts suivants :

- Sont utilisables dans des contextes où les essais randomisés sont difficiles à mener, comme les maladies rares ou les situations urgentes.
- Permettent de répondre à des questions causales en contexte non randomisé, grâce à l'approche d'émulation d'essai clinique.
- Peuvent permettre de réduire les biais (sélection, immortalité, information, confusion sur variables mesurées) par un design rigoureux et des méthodes statistiques adaptées, par rapport à une comparaison naïve de type « benchmark »
- En particulier, un des apports majeurs de l'approche d'émulation d'un bras dans un essai clinique est d'aider à définir le bon « temps zéro », c'est-à-dire l'alignement de critères d'éligibilité validés, de l'assignement des stratégies thérapeutiques et du début de comptabilisation des événements, et ainsi de prévenir les biais de sélection en particulier un biais d'immortalité.

Limites

En pratique, l'utilisation des données issues du soins dans la recherche clinique et notamment pour la constitution de groupes contrôles externes se heurte fréquemment à différents problèmes :

- Impossibilité à ce jour d'identifier des patients similaires à ceux qui seront inclus dans la partie « monobras » (en particulier avec les traitements visant une nouvelle altération moléculaire non mesurée en pratique courante, mais nécessaire pour les thérapies ciblées).
- Indisponibilité des critères de jugements.
- Par exemple, dans des maladies neurologiques, comme la maladie d'Alzheimer, les critères nécessaires à l'évaluation des nouvelles molécules sont des scores d'échelles, p. ex. ADAS-cog (*Alzheimer's Disease Assessment Scale – Cognitive Subscale*) ou autres, non évalués en pratique courante. En oncologie, la progression de la maladie n'est pas évaluée en routine avec des critères standardisés (ex *Response Evaluation Criteria In Solid Tumors* (RECIST)) comme elle peut l'être dans le cadre des essais cliniques.
- Indisponibilité de la date de prise en charge du patient avec la même définition et temporalité que dans l'essai clinique. Afin de pouvoir définir la date index avec un niveau de standardisation suffisant, il faut pouvoir mobiliser des « proxy » comme le temps de la décision de recourir à un traitement.
- La démonstration de l'échangeabilité des patients est difficile en pratique.

De plus, les comparaisons externes mobilisant des données issues du soin courant peuvent présenter des limites concernant de la qualité, l'exactitude, la précision et la complétude des données :

- Indisponibilité de la totalité des facteurs de confusion, qui incluent des variables en dehors du champ de la maladie lorsque le critère de jugement est par exemple la mortalité toutes causes confondues.
- Qualité des données incertaines (précision, exactitude et données manquantes) ne permettant pas d'assurer l'absence d'erreur de mesure symétrique ou asymétrique nécessaire pour la prévention des biais de mesure et de classification, et la prévention du biais d'attrition.
- Pauvreté du suivi longitudinal ne permettant pas de gérer la problématique de la confusion dépendante du temps ou de gérer les événements intercurrents de façon appropriée pour évaluer d'autres «estimands» qu'un «*treatment policy estimand*»(16) .

Exemples d'utilisation

De nombreux exemples existent. On peut citer :

- L'étude CHAMPION-NMOSD – An Efficacy and Safety Study of Ravulizumab in Adult Participants With NMOSD ([NCT04201262](#)) (17–19)
- L'étude LEROS – Study to Assess the Efficacy and Safety of Raxone (idébénone) in LHON Patients ([NCT02774005](#)) (20,21)
- L'étude TOSCA – Cémipimab versus traitements systémiques historiques du carcinome épidermoïde cutané localement avancé ou métastatique (22)

Position des autorités

L'EMA, la FDA et la HAS ont toutes trois publié sur ces essais non contrôlés et réaffirment la place centrale de l'essai randomisé (23–26).

Toutes les recommandations méthodologiques récentes sur ces approches insistent sur le fait qu'elles ne sont pas des alternatives standard à l'ECR, mais bien des solutions d'exception potentiellement utilisables dans des circonstances très particulières.

L'analyse des avis de la commission de la transparence (27) a mis en lumière différentes situations avec des degrés d'acceptabilité assez variables, et repris dans le « position paper » de la HAS (25) :

- Essais non contrôlés avec une comparaison indirecte par contrôle externe : l'essai non contrôlé (« monobras ») est réalisé en premier, puis, ensuite dans un deuxième temps, sont réalisées des comparaisons avec des groupes contrôles externes. Cette 1^{ère} situation peut entraîner un biais, car il s'agit d'une analyse post-hoc, potentiellement influencée par les résultats observés de l'étude, et n'est acceptée que dans des situations particulières et limitées. Une rigueur méthodologique, comme le choix justifié du bras de comparaison externe, qui doit être indépendant des résultats produits, et la préparation de plan d'analyses statistiques en insu des résultats afin d'éviter le « p-hacking » (adaptation de l'analyse statistique en cours de réalisation, en fonction des résultats qu'elle produit (14)), est indispensable pour que la comparaison soit informative.
- Essais contrôlés par un bras de contrôle externe défini *a priori* et présenté aux autorités à l'occasion des rencontres d'amont avec des données collectées en pratique courante ou avec des données issues d'autres études. Cette 2^{ème} situation est davantage considérée par les autorités, dès lors que le bras comparateur externe est défini dès la construction du protocole.

Recommandations et leviers

Il est recommandé de suivre les guidelines (23–26), et en particulier l'approche d'émulation d'essai clinique.

Concernant la qualité des données, elles doivent permettre de :

- Garantir des populations comparables (zone géographique, conditions socio-économiques, modalité de prise en charge...).
- Répondre aux critères d'inclusion de l'étude, *a minima* à la valence « monobras » : pathologies, biomarqueurs, stades, comorbidité
- Identifier une prise en charge thérapeutique similaire aux standards de soins contemporains (standard of care) correspondant aux meilleurs standards de la pratique actuelle (best of care)
- Documenter les critères de jugement nécessaires à la démonstration du bénéfice clinique
- Anticiper le recueil de contrôles négatifs. Ce point pourrait être corrigé en anticipant les ajustements nécessaires à ces études et en enrichissant les sources de données de ces variables (par mise en relation avec d'autres bases ou recueil d'informations supplémentaires, comme le SNDS), et par le recueil de variables pouvant servir de contrôles négatifs ou positifs).
- Extraire la durée de suivi minimale cliniquement pertinente
- Récolter les mesures à l'inclusion des facteurs de confusion et au cours du suivi longitudinal pour les facteurs de confusion dépendant du temps
- Le promoteur de l'étude doit donc connaître les sources de données, leur qualité et leur pertinence a priori.
- En termes de calendrier et pour garantir autant que possible la contemporanéité des données, dès la fin du recrutement prospectif des patients traités avec le nouveau traitement (partie expérimentale de l'étude) ou pour les analyses intermédiaires planifiées, le promoteur demandera l'accès aux données des patients retenus pour le groupe contrôle externe, c'est-à-dire les patients ayant un suivi identique à ceux de l'étude, ceci afin de réaliser les analyses avec la méthodologie appropriée sans retard.

Ces recommandations sur les données étant communes à l'ensemble des méthodologies basées sur l'analyse des données issues du soin courant, elles sont développées dans la section 3.5.

3.3.2. Essais augmentés par données collectées en pratique courante

Définition

Les *externally-augmented trials* (ou essais randomisés augmentés par données externes, parfois appelés *essais hybrides*) consistent à combiner des patients inclus dans un bras contrôle randomisé avec des patients issus d'une source externe (cohorte observationnelle, registre, essai antérieur...). La conception, l'analyse et l'inférence statistique de l'effet du traitement reposent alors sur l'ensemble de ces données. Typiquement, cela permet de réduire le nombre de patients à recruter dans le bras contrôle (par exemple avec un ratio de randomisation 3 :1).

Ici, le terme « hybride » fait référence à l'intégration de contrôles externes en complément d'un bras contrôle randomisé. À noter que ce terme peut aussi désigner d'autres formes hybrides d'essais (ex. : essais décentralisés, c'est-à-dire réalisés en dehors des lieux de recherche classiques comme les hôpitaux ou les cabinets médicaux), non abordées ici.

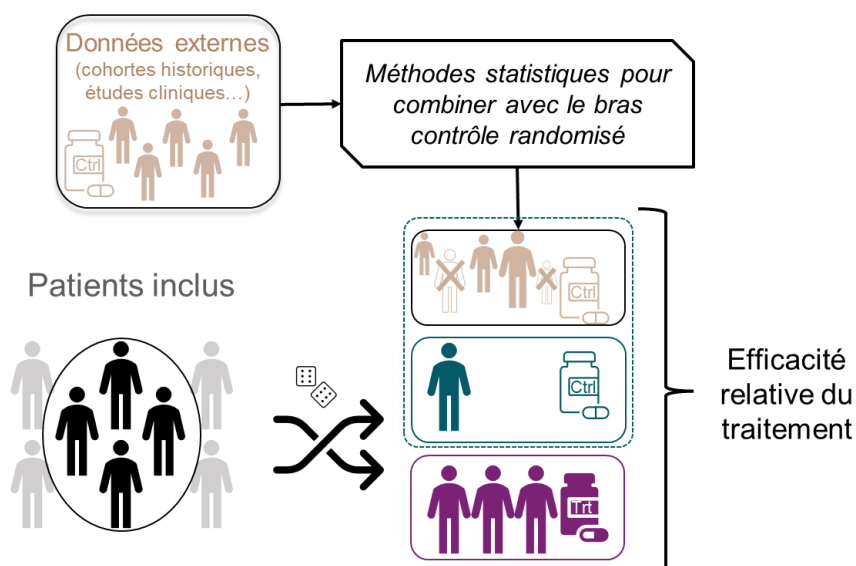


Figure 3 – Illustration des essais « externally-augmented trials »

Plusieurs approches statistiques ont été développées pour estimer les effets du traitement, comme les approches fréquentistes (ou classiques) ou les approches bayésiennes. Des développements méthodologiques (comme les adaptations de randomisation pour inclure les patients du bras externe de façon aléatoire tout en respectant la distribution des covariables, ou une étape d'analyse intermédiaire de vérification de l'hypothèse d'échangeabilité, par exemple), permettent d'améliorer la robustesse de ces approches en les adaptant aux contraintes et aux données spécifiques (28,29).

Intérêt

Cette approche permet de réduire le nombre de patients à recruter dans les bras contrôle. Elle peut également accélérer le développement clinique, limiter l'exposition de patients à des traitements moins efficaces, et tirer parti de données déjà disponibles dans des bases observationnelles ou d'autres essais.

Limites

Les limites concernant les bras comparateurs externes s'appliquent également aux essais augmentés.

L'enjeu majeur réside dans la question de l'échangeabilité entre les patients randomisés et ceux issus des données externes : si les patients de la source externe et les patients de l'essai clinique ne sont pas comparables, l'estimation de l'effet du traitement peut être faussée.

Par ailleurs, cette méthode reste peu courante, principalement à cause des risques de biais. L'expérience pratique reste encore limitée, et la complexité statistique de ces approches peut constituer un frein.

Exemples

L'idée de combiner des contrôles randomisés avec des données historiques ou externes n'est pas nouvelle (30). Cependant, aucun exemple n'a pu être détecté dans les avis de la Commission de la Transparence (CT) de la HAS. Un exemple a été discuté au cours du *Complex Innovative Design* de la FDA (31) sur le lymphome diffus à cellule B.

Acceptabilité par les autorités

L'acceptabilité par les agences de régulation reste conditionnée à une mise en œuvre rigoureuse et transparente. Les autorités attendent que la comparabilité entre groupes soit démontrée et que les méthodes d'analyse soient robustes.

Recommandations et freins à lever

Il est recommandé d'appliquer des méthodes statistiques appropriées (fréquentistes ou bayésiennes) pour limiter les biais potentiels. Des outils comme la randomisation adaptative tenant compte de la distribution des covariables, ou les tests d'échangeabilité à mi-parcours, peuvent améliorer la robustesse de l'analyse.

De plus, des échanges avec les autorités en amont du lancement de l'étude sont fortement recommandés.

Les freins à lever incluent :

- La réticence liée au manque de standardisation de ces approches,
- Le besoin de guides pratiques ou de recommandations claires des autorités,
- Et la nécessité de partager plus largement les retours d'expérience issus d'essais utilisant cette méthodologie.

Il s'agit d'une approche exploratoire, qui nécessite une évaluation et des études de preuves de concept.

Une utilisation responsable et transparente de ces méthodes est indispensable.

3.4. Pour le suivi de l'efficacité en vie réelle

Définition

Les études comparatives non randomisées sont des études observationnelles où le groupe traité et le groupe contrôle sont issus tous les deux des données de vraie vie. Cette approche sous-entend donc que le traitement évalué est utilisé en routine. Ces études peuvent être notamment :

- Les études de cohorte : les patients sont identifiés comme appartenant à un groupe d'étude en fonction du ou des traitements reçus
- Les études cas témoins : les patients sont d'abord identifiés comme appartenant à un groupe d'étude sur la base d'un évènement médical, ... le contexte et les traitements antérieurs reçus sont par la suite identifiés ;
- Les études type « auto-contrôlées » où les patients constituent leur propre témoin.

Par leur nature elles ne sont pas adaptées au développement initial d'un nouveau traitement (qui n'est pas encore utilisé en pratique courante). Elles peuvent cependant être envisagées pour lever l'incertitude sur l'efficacité ou l'impact d'une technologie de santé après l'enregistrement et l'accès au marché.

Ces études doivent être réalisées dans le cadre strict de l'essai émulé, comme décrit en section 3.2. Les experts de l'atelier de Giens et du GT AIS-F CRIN mettent en avant cette méthodologie qualifiée de fiable. Des guidelines spécifiques de données ont aussi été récemment publiées par le consortium PRINCIPLED (9).

Limites

Ces études peuvent être complexes à mettre en œuvre avec de nouveaux traitements considérés très prometteurs, car après sa mise à disposition du soin, le traitement sera utilisé par les médecins pour tous les patients. Ainsi, peu de patients devraient recevoir le traitement de la génération précédente qui devait être le contrôle de l'étude dès la mise à disposition du nouveau traitement. L'inférence causale est donc fragilisée et l'hypothèse de positivité est donc affaiblie. Cette limite est abordée dans le paragraphe sur les recommandations sur les données et le besoin de recueil prospectif et dynamique des données, en section 3.5.

Ces études requièrent un plan d'analyse rigoureux et méticuleux afin de limiter les différents biais inhérents à leur statut d'étude observationnelle, en particulier le biais de confusion, ainsi que la gestion de l'inférence causale, centrale dans cette situation, permettant de conclure correctement sur les résultats de l'étude (14).

Les sources de données posent les mêmes problématiques pour la réalisation de ces études que pour la réalisation des groupes contrôles externes en termes de disponibilité des critères de jugements, des facteurs de confusion potentiels et des contrôles négatifs et positifs. En effet, la réalisation de ces études

requiert, une fois de plus, des données robustes, en particulier en termes de disponibilités 1) des critères de jugement adapté aux études inférentielles de confirmation et qui ne sont pas recueilli dans les systèmes d'information de routine, 2) des facteurs de confusion potentiels et 3) des contrôles négatifs ou positifs.

Position des autorités

La HAS a présenté ses recommandations dans le guide méthodologique (4).

Une très récente guideline de la FDA fixe la méthodologie de ces études (32) et complète les guides déjà disponibles (33–38).

Ces recommandations mettent en avant l'importance de :

- Justifier l'utilisation d'une étude non interventionnelle pour répondre à la question de recherche d'intérêt pour l'indication et le contexte envisagés
- Et de bien pré-spécifier l'ensemble du protocole notamment l'hypothèse et le plan d'analyse statistique en amont de la conduite de l'étude. Les recommandations de la FDA s'articulent ensuite autour du design, des sources de données et du plan d'analyse statistique, renforçant les exigences sur l'ensemble de ces points afin de limiter les biais au même titre que pour les études interventionnelles.

Recommandations et leviers à lever

Pour résumer, le GT recommande de suivre le cadre strict de l'essai émulé (justification, pré-spécification, analyses statistiques rigoureuses, évaluation des biais). Les leviers à actionner concernent l'adéquation et la qualité des données collectées en pratique courante mobilisée (voir les recommandations du GT ci-dessous)

3.5. Recommandations concernant les sources de données collectées en pratique courante

En complément des recommandations formulées dans le précédent rapport du GT AIS-F-CRIN (3), pour optimiser l'implémentation des méthodes développées précédemment avec des sources de données françaises, le GT AIS-F-CRIN formule les préconisations suivantes :

Sensibilisation des porteurs de registres et cohortes à l'utilisation secondaire des données

Il est essentiel de sensibiliser les responsables de registres ou cohortes à la problématique d'utilisation secondaire des données pour la recherche clinique. Cette utilisation requiert une méthodologie différente de celle habituellement employée pour l'exploitation scientifique standard, où l'objectif se limite généralement à l'observation d'une situation particulière et à l'identification de facteurs spécifiques.

Les porteurs de registres ou de cohortes doivent comprendre les conditions requises pour permettre l'utilisation de leurs données à des fins réglementaires liées à l'évaluation des produits de santé.

Adaptation des structures existantes et création de nouvelles sources

Il convient de :

- Créer de nouveaux registres/cohortes adaptés aux besoins spécifiques de la recherche clinique
- Adapter les structures existantes pour garantir la disponibilité des informations indispensables
- Assurer un partage des données recueillies par ces structures

Ces adaptations vont au-delà d'une simple modification de base de données et nécessitent généralement la mise en place d'un recueil renforcé (incluant par exemple des critères de jugement non mesurés en pratique courante).

Modalités de recueil et de suivi

Dans leur domaine de recueil, ces sources de données devront :

- Effectuer un enregistrement prospectif (au fil de l'eau) des patients
- Assurer un suivi longitudinal jusqu'à la survenue des événements recherchés
- Garantir la qualité des données grâce à une organisation structurée des rôles et responsabilités
 - Recueil des données par les Techniciens d'Étude Clinique (TEC)
 - Stockage et gestion par les Data Managers
 - Contrôle qualité par les Attachés de Recherche Clinique (ARC)
 - Émission et suivi des requêtes sur les données manquantes ou incohérentes
 - Monitoring continu de la qualité des données

Les examens paracliniques spécifiques et leur traitement éventuellement nécessaire au recueil ou à la construction du critère de jugement doivent être intégrés de manière prospective (exemple : réponse tumorale et temps jusqu'à progression selon les critères RECIST, recueil systématique du score ECOG).

Gestion des populations ciblées par biomarqueurs

Certaines situations posent des difficultés majeures d'anticipation, notamment le cas des populations ciblées par biomarqueurs (comme les altérations moléculaires en oncologie). L'intérêt des groupes contrôles historiques étant leur disponibilité avant l'administration du nouveau traitement, il est nécessaire d'anticiper la caractérisation des patients avec de nouveaux marqueurs bien avant leur standardisation.

Pour résoudre cette problématique, deux approches sont envisageables :

- Conserver dans les bases de données tout test biomarqueur potentiel réalisé (incluant les informations sur la méthode de mesure)
- Mettre en place une conservation systématique de matériel biologique pour des analyses ultérieures (selon le modèle des biobanques)

Procédures qualité et développement

Des procédures qualité rigoureuses doivent être mises en place. Grâce au recueil prospectif avec suivi et data management, les problématiques d'exactitude des informations seront limitées par rapport aux bases médico-administratives, mais cela représente néanmoins un facteur supplémentaire de complexité technique.

Développements prioritaires et prérequis

Pour faciliter l'usage de cohortes prospectives, il convient de :

- Aligner au mieux les modalités de suivi de la recherche clinique avec la pratique médicale courante (fréquence des consultations/visites, calendrier d'exams d'imagerie pour la survie sans progression)
- Établir un modèle économique pérenne pour le financement des cohortes, couvrant :
 - La phase initiale de constitution des registres et cohortes
 - Le maintien et la gestion à long terme de ces structures (personnel, infrastructure, mise à jour des systèmes)
 - La valorisation de l'exploitation des sources de données
- Élaborer une checklist des exigences méthodologiques mentionnées dans les guides FDA, EMA et HAS afin de démontrer que l'approche répond à chacun de ces critères

Cette dynamique permettra de réaliser des études de validation empirique de l'approche en utilisant ces données pour évaluer des traitements ayant déjà fait l'objet d'essais randomisés contrôlés, et dont les résultats sont connus. Au-delà de leur intérêt scientifique, la publication des résultats de ces études comparatives établira la valeur et la fiabilité de cette approche.

4. Le numérique et l'IA au service de la recherche clinique

4.1. Intérêt des outils numériques pour accélérer et optimiser la conception, la réalisation et l'analyse des essais cliniques.

Comme mentionné précédemment, bien que les essais cliniques soient considérés comme le *gold standard* pour étudier la causalité en médecine, ils se heurtent souvent à des contraintes matérielles et peuvent être difficiles à mettre en place dans certains contextes ou pour certaines pathologies. Les méthodes décrites dans ce chapitre offrent des solutions potentielles pour optimiser ces essais, en permettant de gagner du temps, de réduire les coûts et d'exploiter plus efficacement les ressources disponibles. Trois modèles d'application sont abordés dans ce chapitre :

1. L'usage organisationnel de l'intelligence artificielle, permettant d'optimiser l'efficacité opérationnelle des essais cliniques, de l'identification des participants à la gestion des données. Il convient de noter que les modèles de type grands modèles de langage (*large language models*, LLM), bien que prometteurs dans le domaine de la santé, n'ont pas été inclus dans le périmètre des recommandations présentées dans ce chapitre.
2. L'enrichissement pronostique, permettant d'anticiper l'évolution d'un état de santé, d'une pathologie à partir de la collecte de données et de modéliser la « trajectoire » temporelle du patient à but de traitement, de prédiction et de prévention. Associé à un ajustement sur les covariables pour des variables continues ou des mesures catégorielles, cette méthode peut amplifier le contraste entre bras expérimental et bras contrôle ;
3. Le recours aux modèles dits « *in silico* » qui consistent à appliquer un modèle computationnel, qu'il soit de type mécaniste ou généré à partir des données disponibles (par IA), à des patients virtuels, terme encore en cours de discussion, ou cohortes augmentées par des données artificielles, permettant de compléter les essais cliniques randomisés en apportant des informations (estimées) complémentaires (réponse / non réponse au traitement dans des sous populations, comparaison des prises en charges, etc.).

4.2. Designs et cas d'usage

Applications de l'intelligence artificielle à l'optimisation des essais cliniques

Définition et intérêt

L'IA, dans le cadre des essais cliniques, désigne un ensemble de technologies informatiques capables d'effectuer des tâches habituellement réalisées par des humains, en apprenant à partir de données. Ces technologies incluent des méthodes telles que le traitement automatique des langues (*natural language processing*, NLP), le *machine learning* et les algorithmes d'optimisation. Elles sont utilisées pour automatiser, améliorer ou fiabiliser diverses étapes des essais cliniques : de l'identification des participants à la gestion des données, en passant par le suivi de l'observance ou le recueil de critères de jugement. L'IA permet ainsi de gagner en efficacité, en précision et en qualité, tout en réduisant certains biais ou coûts associés aux démarches classiques.

Exemples d'utilisation

L'IA offre de nombreuses opportunités pour améliorer l'organisation des essais cliniques. Plusieurs usages émergents sont décrits ci-dessous :

1. **Sélection des participants** : L'IA, via des modèles de NLP, peut extraire automatiquement des informations à partir de dossiers médicaux électroniques afin d'identifier les patients éligibles à un essai (39–41). Certains systèmes peuvent même suggérer pour les essais auxquels le profil d'un patient pourrait correspondre. Le développement d'entrepôts de données de santé, notamment hospitaliers, facilite cette approche.
2. **Optimisation de la randomisation** : Des algorithmes d'optimisation peuvent être utilisés pour améliorer les méthodes de randomisation par minimisation via l'IA, en équilibrant mieux les groupes sur des caractéristiques cliniques (42). Cette approche peut augmenter la puissance statistique des essais et permettre de réduire leur taille de 25 à 50 % par rapport aux méthodes classiques.
3. **Suivi de l'observance** : Des outils IA peuvent surveiller l'adhésion des participants au traitement, comme une plateforme qui vérifie la prise de médicament via des selfies (43). Bien que cette technologie ait montré des résultats prometteurs sur le renforcement de l'observance au traitement, elle soulève des questions d'acceptabilité et éthiques potentiellement difficile à reproduire en vie réelle.
4. **Amélioration du recueil de données** : L'IA permet d'exploiter des outils connectés (ex. : accéléromètres) pour collecter des données en continu et enrichir le suivi longitudinal. Elle peut également faciliter l'identification et l'adjudication de critères de jugement à partir de textes médicaux, grâce au NLP (44).
5. **Traitement et contrôle des données** : Enfin, le machine learning peut automatiser certaines tâches comme la vérification de la cohérence des données, l'identification d'erreurs grâce à la lecture optique couplée au NLP, ou encore la gestion des données manquantes.

Limites

Malgré son potentiel, l'utilisation de l'IA dans les essais cliniques présente plusieurs limites :

Limites techniques

- **Dépendance aux données** : l'IA requiert des données nombreuses, de qualité et représentatives pour fonctionner efficacement. Les performances des algorithmes dépendent directement de la qualité et de la quantité des données d'entraînement disponibles.
- **Annotation des données médicales** par l'association d'informations complémentaires (images radiologiques ; dossiers médicaux) qui soient compréhensibles par les algorithmes d'IA.
- **Interopérabilité** : l'hétérogénéité des systèmes d'information de santé et le manque de standardisation des données limitent souvent l'implémentation à grande échelle des solutions d'IA.
- **Complexité d'implémentation** : l'intégration des outils d'IA dans les workflows existants des essais cliniques nécessite des adaptations techniques importantes et une expertise spécifique souvent coûteuse.

Limites éthiques et réglementaires

Protection des données : l'utilisation massive de données personnelles de santé soulève des questions importantes de confidentialité et nécessite des mesures rigoureuses de protection conformes au Règlement Général sur la Protection des Données (RGPD).

- **Transparence** : les modèles d'IA, particulièrement ceux basés sur l'apprentissage profond, peuvent fonctionner comme des "boîtes noires" difficiles à interpréter, ce qui pose problème dans un contexte réglementaire exigeant l'explicabilité et la traçabilité des décisions.

Limites pratiques et d'acceptabilité

- Acculturation des acteurs : des efforts de formation et d'accompagnement des professionnels et des patients seront nécessaires pour permettre l'adoption des technologies d'IA de manière sécurisée.
- Généralisation des résultats : les outils d'IA mobilisés dans la recherche clinique ne sont pas forcément applicables dans la pratique courante, restreignant l'usage du produit de santé développé avec leur recours.

Position des autorités

Au cours des dernières années, les agences réglementaires ont porté une attention croissante à l'impact de l'IA sur le développement, l'évaluation et la surveillance des produits de santé. Face aux enjeux posés par l'intégration de ces technologies dans le cycle de vie du médicament, les régulateurs ont publié des recommandations destinées à encadrer leur utilisation (45).

De façon plus individuelle, les autorités proposent un dialogue précoce avec les porteurs de projets, notamment à travers les procédures de qualification, afin de garantir la robustesse, la transparence et la pertinence de ces outils avant leur mobilisation dans la recherche clinique ou les décisions réglementaires. Jusqu'à récemment, aucun de ces cas d'usage n'avait fait l'objet d'une qualification formelle par l'EMA, les scénarios présentés portant sur des utilisations à faible impact réglementaire. Cette situation a toutefois évolué avec la qualification en mars 2025 par l'EMA du premier outil d'IA destiné à soutenir le diagnostic de la maladie stéato-hépatique métabolique (MASH) à partir d'images de biopsies hépatiques(46). Cette décision marque une étape importante dans la reconnaissance réglementaire de l'IA, et confirme la possibilité de recourir à des modèles algorithmiques pour des usages à fort enjeu clinique, à condition de démontrer leur validité méthodologique et leur applicabilité dans un cadre bien défini.

Modèles pronostics

Définition

Les modèles pronostiques permettent de prédire l'évolution future d'une maladie ou d'une condition clinique chez un individu. Nous présentons ici deux applications pouvant en être faites.

- *Enrichissement pronostique*

L'enrichissement pronostique augmente la puissance d'une étude en ciblant une population de patients plus homogène en termes de risque d'évolution de la maladie. En sélectionnant des patients dont l'évolution est plus prévisible et qui sont plus susceptibles de répondre au traitement, on réduit la variabilité au sein du groupe étudié. Cela permet de mieux détecter les effets du traitement, même avec un nombre fixe de participants.

- *Utilisation des modèles pronostiques pour l'ajustement sur les covariables*

Dans l'analyse de l'efficacité des essais cliniques, des variables d'ajustement sont couramment prises en compte pour corriger les effets confondants causés par des covariables influençant le critère de jugement. Ces variables sont sélectionnées en fonction des connaissances médicales sur la maladie. L'ajustement basé sur des variables pronostiques (c'est-à-dire celles associées à l'issue clinique d'intérêt) améliore la puissance statistique des essais cliniques randomisés. L'intelligence artificielle permet désormais de définir de nouvelles variables pronostiques en utilisant des modalités complexes qui ne sont pas nécessairement accessibles en pratique clinique.

Intérêt

Ces deux applications permettent de mieux contrôler la variabilité au sein de l'essai, d'augmenter la puissance statistique et d'optimiser la conception des études. L'utilisation de l'IA pour le développement de nouvelles variables pronostiques constitue une avancée majeure, notamment dans les cas où les données cliniques traditionnelles ne suffisent pas à capter des éléments complexes.

Limites

L'utilisation de ses deux applications s'accompagnent de pré requis et de limites propres à chacune :

La mise en œuvre de l'enrichissement pronostic nécessite le déploiement du modèle au moment de l'inclusion des patients dans l'essai clinique afin de pouvoir déterminer leur éligibilité.

Puisqu'un essai utilisant cette stratégie prouve l'efficacité du traitement uniquement chez les patients à haut risque de progression, sa mise sur le marché sera limitée à cette catégorie de malades. Pour une application plus large, il sera donc nécessaire soit de déployer le modèle pronostique en tant que test compagnon, soit de mener des essais supplémentaires sur une population plus diversifiée, incluant des patients à risque intermédiaire (47).

L'utilisation de stratégie d'enrichissement à partir de modèles de prédiction innovants implique donc la prise en compte des 3 contraintes suivantes :

- La nécessité d'identifier plus de patients au départ,
- La nécessité de développer un test compagnon pour aller au bout de la démarche,
- Une modification de la population ciblée par l'essai.

En ce qui concerne la gestion du cycle de vie des variables d'ajustement basés sur l'IA, l'utilisation d'une variable d'ajustement dans la stratification de la randomisation peut nécessiter son déploiement au niveau de la procédure d'inclusion d'un essai clinique, déploiement qui peut se faire à partir de la même infrastructure technique et informatique que les outils diagnostiques basés sur l'histologie comme MSIIntuit (48).

Une fois ce préalable posé, il est nécessaire de prendre en compte les éléments suivants :

Lors de la conception de l'essai clinique, une évaluation préalable des performances et un rapport d'évaluation des risques est à réaliser. L'évaluation des performances avant l'essai assure toutes les preuves disponibles en termes de performances et indique si les populations de validation sont représentatives de la population cible de l'essai. L'évaluation des risques liés au modèle, évalue leur probabilité et leur impact et propose des mesures d'atténuation pour réduire leur probabilité d'occurrence.

À la fin de l'essai clinique, une évaluation post-essai évalue les performances du modèle d'IA dans l'essai et également de l'impact des différents risques identifiés dans le rapport préalable à l'essai et l'efficacité des mesures d'atténuation est à réaliser. D'autres questions pertinentes non prévues initialement sont également mises en évidence dans ce rapport.

Un risque identifié est qu'un essai conçu en se basant sur des hypothèses de performance pronostique puisse finalement être sous-dimensionné, entraînant une puissance statistique insuffisante si la performance pronostique est inférieure aux attentes. Pour éviter ce problème, il peut être judicieux de dimensionner l'essai sans intégrer le gain supposé de l'ajustement ou de prévoir une re-estimation en aveugle de la taille de l'échantillon en cours d'étude.

Exemples

- *Enrichissement pronostique*

L'enrichissement sur des biomarqueurs est courant en cancérologie, notamment pour les essais de thérapies ciblées qui restreignent généralement l'inclusion aux patients dont la tumeur présente la mutation ciblée par le traitement (49).

Dans la maladie d'Alzheimer, un modèle utilisant des données multimodales a permis de prédire le déclin cognitif, à partir notamment de l'analyse des déformations de l'hippocampe avec de bien meilleures performances que les modèles existants (50). L'utilisation de ce modèle de progression de la maladie d'Alzheimer qui restreint l'inclusion dans un essai clinique à des patients à haut risque de progression peut alors permettre de réduire la taille d'échantillon nécessaire de 40% par rapport à l'approche actuelle.

- *Utilisation des modèles pronostiques pour l'ajustement sur les covariables*

Une approche appelée PROCOVA a été proposée, reposant à la fois sur des données historiques (groupes témoins d'essais cliniques ou données observationnelles) et des techniques de modélisation pronostique pour améliorer l'analyse des essais cliniques en ajustant le score pronostique. L'avantage de cette

méthode est que l'ajustement sur un score pronostique dans un modèle linéaire (analyse de covariance, ANCOVA) permet de maintenir le risque d'erreur de première espèce, indépendamment de la manière dont le modèle pronostique a été développé. Cela permet notamment d'utiliser des approches de *machine learning* ou d'intelligence artificielle pour développer des modèles avec des performances pronostiques supérieures.

Une autre approche, à ce jour non retenue par les autorités, concerne le développement d'algorithmes de deep learning à partir de lames d'histologie pour la prédiction (51). L'application de ces algorithmes à la prédiction de la survie dans des cohortes de patients atteints de mésothéliome et de carcinome hépatocellulaire réséqués a permis de créer deux modèles pronostiques, MesoNet et HCCnet (52,53). Par exemple, l'ajustement avec HCCnet, en plus des variables classiques, permettrait une réduction de la taille d'échantillon requise de 12 % pour un essai clinique sur un traitement adjuvant pour le carcinome hépatocellulaire réséqué (54).

Position des autorités

Concernant l'enrichissement pronostique, la FDA encadre son utilisation depuis une directive de 2019, avec plusieurs exemples en cardiologie et oncologie (55). Cependant, à ce jour, aucune de ces stratégies n'a encore intégré l'IA ni abouti à un avis formel d'un régulateur. L'approche est néanmoins reconnue comme une option envisageable.

L'enrichissement pronostique peut être combiné à l'ajustement sur des variables innovantes, présenté ci-dessous. À ce sujet, l'EMA (2015) et la FDA (2023) ont publié des directives précisant les conditions d'utilisation de l'ajustement sur covariables (56,57). Celles-ci clarifient plusieurs aspects statistiques et méthodologiques, mais la directive de la FDA exclut l'usage de méthodes d'apprentissage automatique (*machine learning*) pour cet ajustement.

En ce qui concerne les variables d'ajustement innovantes, deux exemples ont fait l'objet d'un avis de l'EMA :

PROCOVA™ (2022) : L'EMA l'a qualifié comme score d'ajustement pronostique, précisant que, bien qu'elle ne recommande pas une méthode spécifique, l'approche présentait des garanties méthodologiques (extension de l'ANCOVA, contrôle du risque d'erreur de première espèce, absence de biais asymptotique) (58).

Mesonet et HCCnet (2023) : L'EMA a reconnu leur potentiel dans les essais cliniques randomisés mais sans qualification officielle (59). Deux réserves ont été émises :

- L'ajustement dans les analyses de survie pourrait accroître le risque d'erreur de première espèce, notamment en cas de risques non proportionnels.
- La généralisation aux essais cliniques multicentriques reste incertaine, compte tenu de la taille limitée des cohortes de validation externes.

En résumé, les autorités reconnaissent l'intérêt de ces approches mais restent prudentes, notamment sur les risques méthodologiques et la généralisation des modèles à grande échelle.

Utilisation de la simulation et du *in silico* dans la recherche clinique

Définition

Les modèles *in silico*, également appelés modèles computationnels, incluent les modèles PKPD, des modèles d'intelligence artificielle, de la modélisation des trajectoires patients et modèles issus des essais pré-clinique biologiques (*Computational Models for Clinical Applications in Personalized Medicine - Guidelines and Recommendations for Data Integration and Model Validation*). Ces méthodes exploitent des bases de données cliniques en utilisant des outils mathématiques et informatiques pour établir des corrélations statistiques entre les informations disponibles et les connaissances existantes.

Il existe plusieurs types d'essais *in silico*, parmi lesquels les approches mécanistes sont les plus avancées. Ces dernières intègrent les connaissances existantes et tirent pleinement parti des ensembles de données disponibles, permettant ainsi une modélisation plus précise et pertinente.

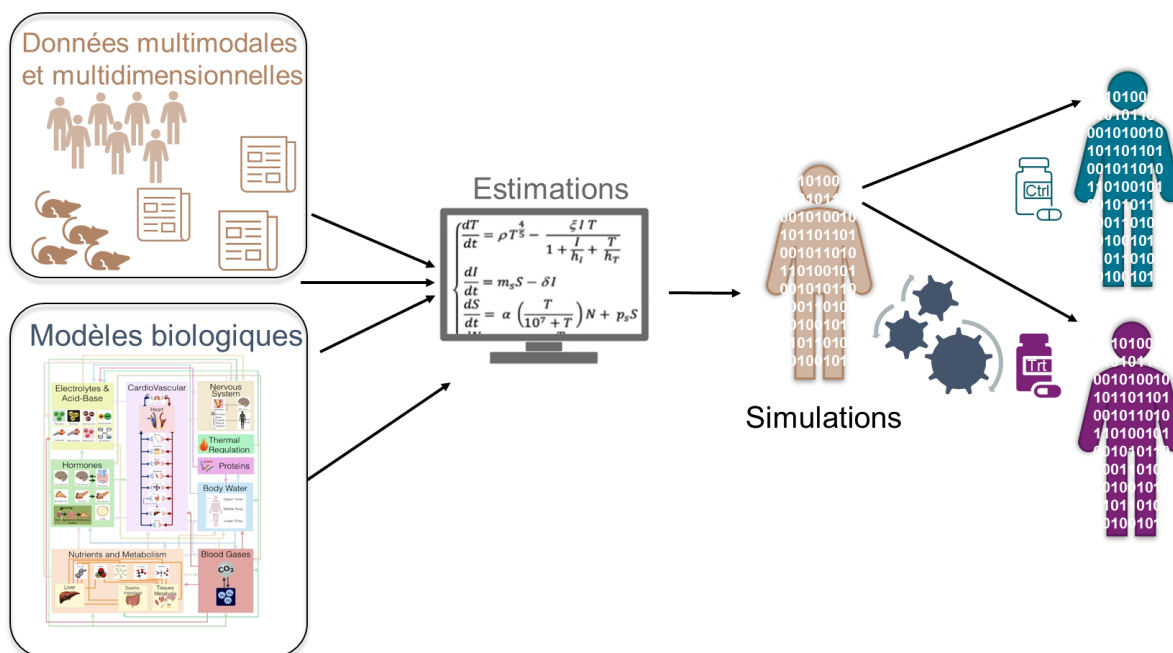


Figure 4 – Illustration des essais in silico (modèles mécanistes)

En complément des méthodes mécanistes, les modèles in silico peuvent provenir de modèle d'IA basés sur l'entraînement des données. Si les modèles mécanistes requièrent une connaissance et compréhension exhaustive des concepts physiopathologiques sous-jacents, cette technique « *data driven* » permet de s'affranchir de cette problématique. Cependant, elle va alors nécessiter un nombre important de données omiques pour entraîner les modèles.

Par exemple, l'utilisation de modèles de *deep learning* tels que les autoencodeurs variationnels (VAEs) permet de générer des données artificielles à partir de données réelles afin de constituer des cohortes augmentées combinant patients réels et patients artificiels.

Cette approche, encore très expérimentale, pourrait permettre de pallier les difficultés de recrutement de certains patients (enfants, femmes enceintes, personnes âgées, patients atteints de maladies rares), de répondre aux besoins des patients refusant de participer aux bras placebo, notamment dans les essais cliniques sur le cancer, et d'amplifier les populations sous-représentées.

Intérêt et limites

Plutôt que de remplacer les essais réels, les essais in silico visent à en renforcer l'efficacité et à enrichir les résultats obtenus. Par exemple, ils permettent d'élargir les conclusions d'une étude à des populations présentant des caractéristiques différentes ou d'inclure des comparateurs supplémentaires. De plus, ils offrent la possibilité d'extrapoler les résultats d'une étude initialement limitée dans le temps afin d'évaluer des effets à plus long terme.

Une approche hybride combinant modélisation et validation expérimentale peut souvent être adoptée pour garantir la robustesse et l'indépendance des prédictions issues des simulations.

Enfin, les essais in silico facilitent l'exploration d'un large éventail d'options et d'hypothèses, contribuant ainsi à l'optimisation de la conception des études expérimentales et à la confirmation des prédictions issues de la modélisation.

La fiabilité des données générées a été démontrée dans plusieurs publications scientifiques, où des comités d'experts en mathématiques, statistiques et médecine ont validé les résultats au moyen de scores spécifiques (50,60,61).

En parallèle des validations scientifiques, l'intégration de bras contrôle augmentés en complément des bras contrôle réels constituera une étape clé pour confirmer la pertinence et l'efficacité des patients artificiels dans les essais cliniques.

Exemples d'utilisation

Devant la diversité des approches de simulation, nous avons réalisé un état de l'art axé sur les principales caractéristiques des études cliniques ayant utilisé des modèles *in silico* dans le cadre de ce rapport. Il en ressort que ces modélisations sont encore peu utilisées ou du moins publiées dans le cadre de la recherche clinique pour la génération de patients virtuels. Les principales utilisations demeurent principalement académiques et centrées sur la recherche amont, avec une adoption encore limitée dans le développement industriel. Ces études restent en interne dans les laboratoires pharmaceutiques et sont davantage présentés dans le cadre de l'accès au marché sans être publiées au préalable. Cette synthèse est présentée en annexe de ce rapport.

Certains outils de simulation ont toutefois été présentés aux agences de régulation dans un objectif de qualification des outils.

Un modèle de progression de la maladie combinée à un simulateur d'essai clinique a été développé à partir de données issues d'essais antérieurs dans le cadre de la maladie d'Alzheimer (62). Ce modèle permet de prédire l'évolution d'un score (le CDR-SB) chez les patients atteints de formes légères à modérées de la maladie, et d'évaluer l'impact de différentes conceptions d'essais cliniques (durée, taille d'échantillon, critères d'inclusion) sur la puissance statistique ou la détection d'un effet traitement. Il est utilisé en amont des essais cliniques de phase III pour affiner les hypothèses de conception, identifier des critères d'éligibilité plus performants, et justifier certaines décisions méthodologiques dans les plans de développement.

À l'inverse, plusieurs modèles de simulation encore en phase exploratoire ont été proposés pour d'autres pathologies, comme la dystrophie musculaire de Duchenne (63). Le modèle DMD-CTSP (Clinical Trial Simulation Platform) repose sur une modélisation pharmacométrique de la progression de la fonction motrice (6-minute walk test) chez les patients atteints de DMD, intégrant les effets de l'âge, du génotype et des traitements standards. Il est principalement destiné à être utilisé en amont de la conception des essais, notamment pour tester virtuellement différents scénarios dans des populations rares ou difficilement accessibles, comme les enfants non ambulatoires. Toutefois, en l'absence de qualification réglementaire, son intégration dans les soumissions réglementaires reste limitée à des discussions exploratoires.

Une troisième utilisation de la simulation concerne la reproduction d'un essai clinique virtuel à partir d'un modèle mécaniste de la maladie (64). Dans le cadre de la prévention secondaire des maladies cardiovasculaires, une modélisation a été développée pour simuler l'effet d'un traitement hypolipémiant sur la survenue d'événements cliniques majeurs. Le modèle, basé sur des équations différentielles décrivant la progression de l'athérosclérose et calibré à partir d'essais antérieurs, a permis de générer une population virtuelle réaliste et de simuler un essai complet (étude SIRIUS). L'objectif était d'anticiper l'effet à long terme d'une réduction du LDL-cholestérol sur le risque cardiovasculaire. Ce type d'essai simulé, bien qu'innovant, n'a pas encore reçu de validation réglementaire.

Position des autorités

Les autorités sont confrontées à ces simulations depuis plusieurs années déjà avec le développement des extrapolations. Ainsi, l'EMA a établi une stratégie d'évaluation de ces modèles basée sur le risque et sur l'impact réglementaire (65).

La simulation a ainsi pu être mobilisée pour accompagner le développement et la génération de la preuve en pédiatrie, en repartant des données adultes et, le cas échéant, de données pédiatriques issues d'autres indications pour un même produit.

L'EMA a publié en 2023 un avis de qualification favorable pour le modèle de simulation développé dans la maladie d'Alzheimer. Cet avis souligne la robustesse de la démarche méthodologique (validation croisée, performance prédictive sur données externes, prise en compte des incertitudes), ainsi que la pertinence du modèle pour la planification d'essais de phase III. L'agence a reconnu que l'utilisation du modèle permettait de justifier certains choix de conception (critères d'inclusion, taille d'échantillon, durée de suivi), d'optimiser la puissance statistique et de soutenir des stratégies d'enrichissement ou d'analyses intermédiaires. Ce positionnement opérationnel du modèle, comme outil d'aide à la conception et à la justification dans les soumissions réglementaires, a été un élément déterminant dans la décision de qualification.

En revanche, les modèles tels que le DMD-CTSP n'ont à ce jour reçu qu'un soutien formel sous forme de « letter of support » (63). L'EMA a salué l'intérêt potentiel de la plateforme pour améliorer la faisabilité des essais pédiatriques dans la DMD, notamment en population rare ou vulnérable. Toutefois, plusieurs axes d'amélioration en vue d'une qualification ont été identifiées : données de calibration issues de cohortes restreintes, incertitudes sur la variabilité interindividuelle, et manque de validation externe robuste. De plus, l'usage du modèle est resté jusqu'ici limité à des phases exploratoires, sans intégration directe dans des protocoles cliniques ou des dossiers réglementaires. Cette absence d'application prospective, combinée aux incertitudes méthodologiques, explique que le modèle n'ait pas encore atteint le niveau de maturité requis pour une qualification.

Des outils de simulation ont été validés par les autorités de santé mais à ce jour, aucun essai comprenant tout un bras simulé n'a été présenté à la FDA ou l'EMA pour qualification.

L'essai complet simulé n'a pas été présenté pour qualification auprès des agences réglementaires mais a été soumis à la Haute Autorité de santé dans le cadre d'un accès au marché. Cette dernière a reconnu l'originalité de cette approche, ainsi que son intérêt potentiel dans des contextes où les données cliniques sont encore limitées. Toutefois, plusieurs réserves ont été formulées : absence de validation prospective indépendante, transparence insuffisante sur certaines hypothèses clés du modèle, et manque d'articulation claire avec un plan de développement fondé sur des essais réels. En l'absence de procédure de qualification formelle ou d'usage prospectif dans une soumission réglementaire, l'essai simulé a été considéré comme une source exploratoire d'information, sans valeur démonstrative pour l'évaluation du bénéfice clinique (64).

En résumé, les autorités reconnaissent l'intérêt croissant des modèles de simulation dans le développement clinique et encouragent leur utilisation encadrée, en particulier dans des contextes où les données empiriques sont limitées. Toutefois, elles maintiennent des exigences élevées en matière de transparence, de documentation des hypothèses, de validation méthodologique et d'articulation avec des stratégies de développement fondées sur des données probantes. L'acceptabilité réglementaire de ces modèles dépend étroitement de leur degré de maturité scientifique, de la clarté de leurs objectifs d'utilisation et de leur intégration dans un cadre prospectif cohérent. Dans cette perspective, l'EMA a annoncé la publication prochaine de recommandation dédiée à l'évaluation et au reporting des modèles mécanistes utilisés dans le cadre du *model-informed drug development* (MIDD), ainsi que l'organisation d'un workshop avec l'ensemble des parties prenantes pour discuter des critères de qualification et des attentes méthodologiques associées. Ces initiatives confirment la volonté des régulateurs de structurer un cadre d'évaluation partagé, fondé sur la rigueur scientifique et l'harmonisation des pratiques.

Sur la base de ces orientations, il est recommandé d'intégrer dès les premières étapes du développement des modèles *in silico* une stratégie de génération de données et une vision claire de leur usage réglementaire. Cette anticipation est particulièrement déterminante pour les modèles à fort impact potentiel sur la décision, que ce soit dans le cadre d'une soumission à l'EMA ou d'une évaluation médico-économique par la HAS. L'alignement précoce des objectifs scientifiques et réglementaires constitue un levier clé pour renforcer la crédibilité et l'utilisabilité de ces approches dans les processus décisionnels.

Recommandations pour favoriser la diffusion du modèle « In silico »

Reco 1 – Créer une plateforme collaborative pré-compétitive, regroupant les acteurs académiques, industriels et les sociétés savantes, visant à développer et formaliser une bibliothèque de modèles validés. Ces modèles, validés par des experts et autorités sanitaires, seraient ensuite accessibles à tous les industriels, qui pourraient les implémenter et les adapter selon leurs besoins et leur contexte spécifique.

Une première étape pour engager la collaboration pourrait être de créer une plateforme collaborative pré-compétitive de données/cohortes simulées non re-identifiables pour la recherche méthodologique, médicale et en santé numérique

Reco 2 – Mettre en place une validation hybride systématique pour les modèles in silico, alliant à la fois modélisation et validation clinique prospective lorsque l'outil est destiné à accompagner et/ou soutenir un développement clinique. Ce processus permettrait de tester les hypothèses du modèle dans des contextes réels tout en validant l'efficacité des simulations à travers des essais cliniques expérimentaux. La combinaison de simulations et de validation clinique permet de renforcer la crédibilité des prédictions des modèles.

Reco 3 – Lors de la soumission d'un protocole d'essai clinique à des fins réglementaires, prévoir explicitement l'utilisation de données simulées, en définissant leur rôle dans le protocole et dans le plan d'analyse statistique. Cette anticipation permet d'assurer la transparence, la confiance dans les prédictions du modèle et d'encadrer l'exploitation de ces données dans l'interprétation des résultats. Il est fortement recommandé d'y associer des études de sensibilité, afin de tester la robustesse des conclusions aux variations des hypothèses du modèle, et de quantifier l'incertitude liée aux prédictions.

Reco 4 – Cartographier et poursuivre les efforts de développement d'une offre de formation initiale et continue et d'acculturation pour les chercheurs, régulateurs, et industriels sur l'IA en santé et plus particulièrement les outils de modélisation in silico, afin d'assurer une disponibilité des compétences ainsi qu'une compréhension commune des méthodologies et des processus d'approbation réglementaire associés. Cela peut se faire via mise en place d'unités mixtes de services ou de structures privées du même type pour assurer du service

Reco 5 – Mise en place de cas d'études concrets, collaboratifs, impliquant des acteurs académiques, industriels et autorités sanitaires, visant à tester et valider les modèles in silico dans des contextes cliniques réels. Ces cas d'études devraient combiner modélisation in silico et validation clinique hybride, intégrant à la fois des données réelles et simulées pour confirmer la validité des modèles dans des essais cliniques prospectifs.

Les cas d'études permettent de tester l'efficacité des modèles dans des situations concrètes et d'identifier les éventuelles limitations ou améliorations nécessaires. Ce processus accélérera la validation des modèles hybrides et cliniques, fournissant des preuves robustes et des données.

5. Conclusion

Ce rapport propose un panorama de méthodologies innovantes, dont les niveaux de maturité restent encore hétérogènes.

Le cadre entourant l'utilisation des données et l'émulation partielle (bras de contrôle externe) ou totale (études observationnelles) des essais cliniques apparaît aujourd'hui mieux défini. Ces approches ont bénéficié de nombreuses preuves ces dernières années, et leur reconnaissance méthodologique s'est consolidée.

À ce jour, ces méthodes sont principalement mobilisées pour l'optimisation du design des études ou dans le cadre d'analyses post-autorisation de mise sur le marché (post-AMM). Leur utilisation dans des études pivotales de Phase III, en vue de l'enregistrement, reste encore limitée. Des travaux de validation conséquents seront nécessaires pour permettre leur acceptation par les autorités réglementaires dans ce contexte.

En ce qui concerne les essais *in silico* et l'intégration de l'intelligence artificielle en recherche clinique, une validation rigoureuse reste essentielle. Au-delà de la performance technique des outils, il sera nécessaire de démontrer leur réelle valeur ajoutée en pratique clinique.

Nous recommandons une démarche en trois étapes pour valider cliniquement ces innovations :

1. Validation sur un large panel de données existantes
2. Intégration en add-on dans des essais prospectifs
3. Suivi dans le cadre d'études longitudinales mobilisant données collectées en pratique courante et IA

À chaque étape, la transparence, la possibilité d'audit et la coopération entre toutes les parties prenantes seront déterminantes pour instaurer la confiance et garantir l'adoption durable de ces outils.

6. Bibliographie

1. Mire MJ de. Presse - Ministère des Finances. 2023 [cité 7 avr 2025]. 29/08/2023 - Régulation des produits de santé. Disponible sur: <https://presse.economie.gouv.fr/29082023-regulation-des-produits-de-sante/>
2. BASCHET L, DUTEIL E. L'évolution des méthodologies d'essais cliniques : nouveaux outils, nouveaux usages et conditions de recours Identification et analyse de cas dans les évaluations par les organismes régulateurs (FDA, EMA) et l'autorité évaluatrice (HAS [Internet]. 2024. Disponible sur: <https://www.fcrin.org/sites/default/files/brique/fichier/06-2024/Evolution%20des%20m%C3%A9thodologies%20d%27essais%20cliniques%20-%20cas%20-%20Rapport%20-%20F-CRIN%20V3.0.pdf>
3. GT AIS/F-CRIN. Pour l'utilisation des données de santé en vie réelle en recherche clinique Etat des lieux, analyse et recommandations [Internet]. 2024. Disponible sur: https://www.fcrin.org/sites/default/files/brique/fichier/05-2024/Rapport_Pour%20l%27E2%80%99utilisation%20des%20donn%C3%A9es%20de%20sant%C3%A9%20en%20vie%20r%C3%A9elle%20en%20recherche%20clinique_16%20mai%202024_0.pdf
4. Haute Autorité de Santé. Guide méthodologique « Etudes en vie réelle pour l'évaluation des médicaments et dispositifs médicaux » [Internet]. 2021. Disponible sur: https://www.has-sante.fr/upload/docs/application/pdf/2021-06/guide_etude_en_vie_reelle_medicaments__dm.pdf
5. Concato J, Corrigan-Curay J. Real-World Evidence — Where Are We Now? *New England Journal of Medicine*. mai 2022;386(18):1680-2.
6. Wang SV, Schneeweiss S, RCT-DUPLICATE Initiative, Franklin JM, Desai RJ, Feldman W, et al. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials. *JAMA*. avr 2023;329(16):1376.

7. Hernán MA, Dahabreh IJ, Dickerman BA, Swanson SA. The Target Trial Framework for Causal Inference From Observational Data: Why and When Is It Helpful? *Ann Intern Med.* 18 févr 2025;ANNALS-24-01871.
8. FDA. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry [Internet]. 2024. Disponible sur: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory>
9. Desai RJ, Wang SV, Sreedhara SK, Zobotka L, Khosrow-Khavar F, Nelson JC, et al. Process guide for inferential studies using healthcare data from routine clinical practice to evaluate causal effects of drugs (PRINCIPLED): considerations from the FDA Sentinel Innovation Center. *BMJ.* 12 févr 2024;e076460.
10. Esnault C, Baschet L, Barbet V, Chenuc G, Pérol M, Thokagevistk K, et al. Addressing challenges with Matching-Adjusted Indirect Comparisons to demonstrate the comparative effectiveness of entrectinib in metastatic ROS-1 positive Non-Small Cell Lung Cancer. *BMC Med Res Methodol.* 1 mars 2025;25(1):57.
11. Popat S, Liu SV, Scheuer N, Hsu GG, Lockhart A, Ramagopalan SV, et al. Addressing challenges with real-world synthetic control arms to demonstrate the comparative effectiveness of Pralsetinib in non-small cell lung cancer. *Nat Commun.* 17 juin 2022;13(1):3500.
12. Antoine A, Pérol D, Robain M, Delalogue S, Lasset C, Drouet Y. Target trial emulation to assess real-world efficacy in the Epidemiological Strategy and Medical Economics metastatic breast cancer cohort. *J Natl Cancer Inst.* 8 août 2023;115(8):971-80.
13. Antoine A, Pérol D, Robain M, Bachelot T, Choquet R, Jacot W, et al. Assessing the real-world effectiveness of 8 major metastatic breast cancer drugs using target trial emulation. *European Journal of Cancer.* déc 2024;213:115072.
14. SFPT. De la nécessité de la méthodologie dans l'évaluation des médicaments. 2022.
15. Cucherat M, Demarcq O, Chassany O, Le Jeune C, Borget I, Collignon C, et al. Attentes méthodologiques pour la démonstration de l'efficacité des produits de santé par les études observationnelles. *Thérapies.* oct 2024;S0040595724001689.
16. Sun Y, McCulloch CE, Marr KA, Huang CY. Recurrent Events Analysis With Data Collected at Informative Clinical Visits in Electronic Health Records [Internet]. arXiv; 2020 [cité 7 avr 2025]. Disponible sur: <http://arxiv.org/abs/2004.12234>
17. CHMP. Assessment report Ultomiris International non-proprietary name: ravulizumab Procedure No. EMEA/H/C/004954/II/0032 [Internet]. 2023. Disponible sur: https://www.ema.europa.eu/en/documents/variation-report/ultomiris-h-c-004954-ii-0032-epar-assessment-report-variation_en.pdf
18. Avis de transparence ravulizumab ULTOMIRIS 100 mg/ml extension d'indication [Internet]. 2023. Disponible sur: https://www.has-sante.fr/upload/docs/evamed/CT-20332_ULTOMIRIS_PIC_EI_AvisDef_CT20332_EPI789.pdf
19. Kemp A. Update on US regulatory review of Ultomiris in NMOSD [Internet]. 2023. Disponible sur: <https://www.astrazeneca.com/media-centre/press-releases/2023/update-on-us-review-of-ultomiris-for-nmosd.html>
20. EMA. EPAR Raxone Summary of product characteristics [Internet]. 2024. Disponible sur: https://www.ema.europa.eu/en/documents/product-information/raxone-epar-product-information_en.pdf
21. HAS. Avis de transparence idébénone RAXONE 150 mg, comprimés pelliculés Réévaluation [Internet]. 2022. Disponible sur: https://www.has-sante.fr/upload/docs/evamed/CT-19245_RAXONE_PIC_REEVAL_INS_AvisD%C3%A9f_CT19245.pdf
22. HAS. Avis de transparence cémiplimab LIBTAYO 350 mg, solution à diluer pour perfusion Réévaluation à la demande du laboratoire [Internet]. 2023. Disponible sur: https://www.has-sante.fr/upload/docs/evamed/CT-20301_LIBTAYO_PIC_REEV_AvisDef_CT20301.pdf
23. Establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing authorisation | European Medicines Agency (EMA) [Internet]. 2024 [cité 1 oct 2024]. Disponible sur: <https://www.ema.europa.eu/en/establishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketing-authorisation>
24. FDA. Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products (Draft guidance). févr 2023; Disponible sur: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products>
25. Vanier A, Fernandez J, Kelley S, Alter L, Semenzato P, Alberti C, et al. Rapid access to innovative medicinal products while ensuring relevant health technology assessment. Position of the French National Authority for Health. *BMJ EBM.* févr 2024;29(1):1-5.

26. EMA. Reflection paper on establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing authorisation. 17 avr 2023;
27. Quelles évolutions méthodologiques en recherche clinique ? Partage des premiers travaux du groupe de travail mis en place par l'Agence de l'innovation en santé et l'infrastructure de recherche clinique F-CRIN | F-CRIN [Internet]. [cité 7 avr 2025]. Disponible sur: <https://www.fcrin.org/espace-media/actualites/quelles-evolutions-methodologiques-en-recherche-clinique-partage-des>
28. Kanamori S, Takeuchi M. Treatment effect estimation using the propensity score in clinical trials with historical control. *BMC Medical Research Methodology*. févr 2024;24(1):47.
29. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*. 2014;13(1):41-54.
30. Pocock SJ. The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*. mars 1976;29(3):175-88.
31. Research C for DE and. Complex Innovative Trial Design Meeting Program. FDA [Internet]. 8 sept 2024 [cité 7 avr 2025]; Disponible sur: <https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program>
32. FDA, CDER. Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products Guidance for Industry. mars 2024;
33. Berger ML, Mamdani M, Atkins D, Johnson ML. Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I. *Value in Health*. nov 2009;12(8):1044-52.
34. Wang SV, Pottegård A, Crown W, Arlett P, Ashcroft DM, Benchimol EI, et al. HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force. *Pharmacoepidemiology and Drug Safety*. janv 2023;32(1):44-55.
35. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler H, Goettsch W, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf*. sept 2017;26(9):1033-9.
36. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2009;12(8):1053-61.
37. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective Observational Studies to Assess Comparative Effectiveness: The ISPOR Good Research Practices Task Force Report. *Value in Health*. mars 2012;15(2):217-30.
38. Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ*. janv 2021;m4856.
39. Zhang K, Demner-Fushman D. Automated classification of eligibility criteria in clinical trials to facilitate patient-trial matching for specific patient populations. *J Am Med Inform Assoc*. 1 juill 2017;24(4):781-7.
40. Kim J, Quintana Y. Review of the Performance Metrics for Natural Language Systems for Clinical Trials Matching. *Stud Health Technol Inform*. 6 juin 2022;290:641-4.
41. Hassanzadeh H, Karimi S, Nguyen A. Matching patients to clinical trials using semantically enriched document representation. *J Biomed Inform*. mai 2020;105:103406.
42. Bertsimas D, Johnson M, Kallus N. The Power of Optimization Over Randomization in Designing Experiments Involving Small Samples. *Operations Research*. août 2015;63(4):868-76.
43. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using Artificial Intelligence to Reduce the Risk of Nonadherence in Patients on Anticoagulation Therapy. *Stroke*. mai 2017;48(5):1416-9.
44. Cunningham JW, Singh P, Reeder C, Claggett B, Marti-Castellote PM, Lau ES, et al. Natural Language Processing for Adjudication of Heart Failure in a Multicenter Clinical Trial: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Cardiol*. 1 févr 2024;9(2):174-81.
45. EMA. The use of Artificial Intelligence (AI) in the medicinal product lifecycle | European Medicines Agency (EMA) [Internet]. 2024 [cité 6 avr 2025]. Disponible sur: <https://www.ema.europa.eu/en/use-artificial-intelligence-ai-medicinal-product-lifecycle>

46. EMA. EMA qualifies first artificial intelligence tool to diagnose inflammatory liver disease (MASH) in biopsy samples [Internet]. 2025 mars. Disponible sur: <https://www.ema.europa.eu/en/news/ema-qualifies-first-artificial-intelligence-tool-diagnose-inflammatory-liver-disease-mash-biopsy-samples>
47. EMA. Haute Autorité de Santé. [cité 6 avr 2025]. Thérapies ciblées et tests compagnons : méthodes d'évaluation. Disponible sur: https://www.has-sante.fr/jcms/c_1734882/fr/therapies-ciblees-et-tests-compagnons-methodes-d-evaluation
48. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol*. 1 août 2019;30(8):1232-43.
49. Vivot A, Li J, Zeitoun JD, Mourah S, Crequit P, Ravaud P, et al. Pharmacogenomic biomarkers as inclusion criteria in clinical trials of oncology-targeted drugs: a mapping of ClinicalTrials.gov. *Genet Med*. août 2016;18(8):796-805.
50. Koval I, Bône A, Louis M, Lartigue T, Bottani S, Marcoux A, et al. AD Course Map charts Alzheimer's disease progression. *Sci Rep*. 13 avr 2021;11(1):8020.
51. Fu Y, Karanian M, Perret R, Camara A, Le Loarer F, Jean-Denis M, et al. Deep learning predicts patients outcome and mutations from digitized histology slides in gastrointestinal stromal tumor. *NPJ Precis Oncol*. 24 juill 2023;7(1):71.
52. Courtiol P, Maussion C, Moarii M, Pronier E, Pilcer S, Sefta M, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat Med*. oct 2019;25(10):1519-25.
53. Liu QP, Xu X, Zhu FP, Zhang YD, Liu XS. Prediction of prognostic risk factors in hepatocellular carcinoma with transarterial chemoembolization using multi-modal multi-task deep learning. *EClinicalMedicine*. 7 juin 2020;23:100379.
54. Ding XX, Zhu QG, Zhang SM, Guan L, Li T, Zhang L, et al. Precision medicine for hepatocellular carcinoma: driver mutations and targeted therapy. *Oncotarget*. 6 juin 2017;8(33):55715-30.
55. Research C for DE and. Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products [Internet]. FDA; 2019 [cité 6 avr 2025]. Disponible sur: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enrichment-strategies-clinical-trials-support-approval-human-drugs-and-biological-products>
56. EMA. Adjustment for baseline covariates in clinical trials - Scientific guideline | European Medicines Agency (EMA) [Internet]. 2015 [cité 6 avr 2025]. Disponible sur: <https://www.ema.europa.eu/en/adjustment-baseline-covariates-clinical-trials-scientific-guideline>
57. FDA C for DE and. Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products [Internet]. FDA; 2024 [cité 6 avr 2025]. Disponible sur: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products>
58. EMA. Qualification opinion for Prognostic Covariate Adjustment (PROCOVA™) [Internet]. EMA; 2022 sept. Disponible sur: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-prognostic-covariate-adjustment-procovatm_en.pdf
59. EMA. Letter of Support for the statistical adjustment on deep learning prognosis covariates obtained from histological slides [Internet]. EMA; 2023 avr. Disponible sur: https://www.ema.europa.eu/en/documents/other/letter-support-statistical-adjustment-deep-learning-prognosis-covariates-obtained-histological-slides_en.pdf
60. Maheux E, Koval I, Ortholand J, Birkenbihl C, Archetti D, Bouteloup V, et al. Forecasting individual progression trajectories in Alzheimer's disease. *Nat Commun*. 10 févr 2023;14(1):761.
61. Chadebec C, Thibeau-Sutre E, Burgos N, Allassonnière S. Data Augmentation in High Dimensional Low Sample Size Setting Using a Geometry-Based Variational Autoencoder. *IEEE Trans Pattern Anal Mach Intell*. 1 mars 2023;45(3):2879-96.
62. EMA. Qualification opinion of a novel data driven model of disease progression and trial evaluation in mild and moderate Alzheimer's disease [Internet]. EMA; 2013 sept. Disponible sur: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-novel-data-driven-model-disease-progression-and-trial-evaluation-mild-and-moderate-alzheimers-disease_en.pdf
63. EMA. Letter of Support of Model-based Clinical Trial Simulation Platform (CTSP) for Duchenne Muscular Dystrophy [Internet]. EMA; 2022 nov. Disponible sur: https://www.ema.europa.eu/en/documents/other/letter-support-model-based-clinical-trial-simulation-platform-ctsp-duchenne-muscular-dystrophy_en.pdf
64. HAS. Haute Autorité de Santé. [cité 15 avr 2025]. LEQVIO (inclisiran) - Dyslipidémie. Disponible sur: https://www.has-sante.fr/jcms/p_3538269/fr/leqvio-inclisiran-dyslipidemie

65. EMA. Extrapolation of efficacy and safety in paediatric medicine development - Scientific guideline | European Medicines Agency (EMA) [Internet]. 2018 [cité 15 avr 2025]. Disponible sur: <https://www.ema.europa.eu/en/extrapolation-efficacy-safety-paediatric-medicine-development-scientific-guideline>
66. Drysdale SB, Cathie K, Flamein F, Knuf M, Collins AM, Hill HC, et al. Nirsevimab for Prevention of Hospitalizations Due to RSV in Infants. *N Engl J Med*. 28 déc 2023;389(26):2425-35.
67. Haute Autorité de Santé [Internet]. [cité 7 avr 2025]. BEYFORTUS (nir sévimab) - Virus respiratoire syncytial. Disponible sur: https://www.has-sante.fr/jcms/p_3456503/fr/beyfortus-nirsevimab-virus-respiratoire-syncytial
68. iqvia_efpia_pipeline-review_final.pdf [Internet]. [cité 27 sept 2024]. Disponible sur: https://www.efpia.eu/media/602564/iqvia_efpia_pipeline-review_final.pdf

