

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* »

Annexes
Mai 2025

Table des matières

Annexe 1. Glossaire et acronymes _____	2
Annexe 2 Enjeux règlementaires de l'usage de méthodologies complémentaires et/ou alternatives aux essais cliniques ____	6
Annexe 3 - Cross validation des solutions proposées avec les recommandations émises par EMA sur les essais non contrôlés « monobras » _____	25
Annexe 4 - Analyse des recommandations internationales sur la simulation in silico et l'intelligence artificielle _____	28

Annexe 1. Glossaire et acronymes

GLOSSAIRE

Un glossaire a été rédigé conjointement avec l'Alliance SILICA, association pour l'étude et le suivi de l'apport des Données Artificielles en Santé (<https://silica-asso.fr/glossaire-silica/>). Les définitions ci-après en sont un extrait et reflètent le consensus des discussions du collège d'experts qui en sont membres.

Bras de contrôle ou Groupe contrôle : Désigne le groupe de patients ne recevant pas le traitement en cours d'évaluation qui apporte le « contrefait » du groupe traité afin d'isoler l'effet propre du traitement (ref . https://database.ich.org/sites/default/files/E10_Guideline.pdf)

Bras de contrôle (ou groupe contrôle) externe : bras de contrôle (parfois appelé synthétique) utilisant tout type de données non générées au cours de l'essai clinique, par exemple issues des soins ou d'autres études (observationnelles, essais cliniques antérieurs) –

(Ref https://database.ich.org/sites/default/files/E10_Guideline.pdf)

Bras simulé : bras obtenu par simulation numérique résultant de l'exécution d'un modèle du groupe de patients, donc composé de *patients virtuels*. NB: parfois appelé bras de contrôle artificiel. A noter également l'utilisation de la notion de bras augmenté qui associe des données recueillies au cours de l'essai clinique et des données externes soit issues du soin, d'essais cliniques antérieurs, de bases de données existantes ou virtuelles.

Cohorte : Groupe de personnes partageant une caractéristique commune (maladie, traitement, exposition à un facteur de risque) et suivies longitudinalement sur une période pour observer l'évolution de leur santé ou l'apparition de certains événements (guérison, complication, décès...).

Patient virtuel : *patient issu d'une modélisation* qui représente les caractéristiques particulières d'un individu-type dont le profil anatomique, le mode de vie et autres caractéristiques sont pré-définis, afin de soutenir une prise de décision clinique individualisée.

Population virtuelle : ensemble de patients virtuels permettant d'obtenir un sous-ensemble de population plausible en reflétant la variabilité inter-individuelle des patients.

Les essais et études cliniques

Déploieur : « une personne physique ou morale, une autorité publique, une agence ou un autre organisme utilisant sous sa propre autorité un système d'IA sauf lorsque ce système est utilisé dans le cadre d'une activité personnelle à caractère non professionnel » (article 3 du règlement sur l'IA).

Détenteur de données : « toute personne physique ou morale, autorité publique, agence ou autre organisme dans les secteurs des soins de santé ou des soins, y compris les services de remboursement si nécessaire, ainsi que toute personne physique ou morale qui développe des produits ou des services destinés aux secteurs de la santé, des soins de santé ou des soins, qui développe ou produit des applications de bien-être, qui effectue des travaux de recherche ayant trait aux secteurs des soins de santé ou des soins, ou qui agit en tant que registre de mortalité, ainsi que toute institution ou tout organe ou organisme de l'Union européenne, qui ont :

- Le droit ou l'obligation, conformément au droit de l'Union européenne ou au droit national applicables et en leur qualité de responsable ou de responsable conjoint du traitement, de traiter des données de santé électroniques à caractère personnel à des fins de fourniture de soins de santé ou de soins, ou à des fins de santé publique, de remboursement, de recherche, d'innovation, d'élaboration des politiques, de statistiques officielles ou de sécurité des patients ou à des fins de réglementation; ou

- La capacité de mettre à disposition des données de santé électroniques à caractère non personnel par le biais du contrôle de la conception technique d'un produit et de services liés, y compris en enregistrant ou en fournissant ces données, en limitant l'accès à ces données ou en échangeant ces données » (article 2 du règlement EEDS).

Donnée à caractère personnel : toute information se rapportant à une personne physique identifiée ou identifiable ». Est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale. (article 4 du RGPD).

Fournisseur : « une personne physique ou morale, une autorité publique, une agence ou tout autre organisme qui développe ou fait développer un système d'IA ou un modèle d'IA à usage général et le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit » (article 3 du Règlement européen sur l'IA du 13 juin 2024).

Essai clinique monobras : étude interventionnelle non contrôlée et donc non randomisée. Un bras de contrôle externe peut être associé à cet essai.

Etude observationnelle : Etude non interventionnelle, qui peut utiliser des données collectées en pratique courante. Elles peuvent être comparatives entre plusieurs types de prise en charge.

Emulation d'essai clinique : l'émulation d'essai consiste à aligner les questions causales de l'étude observationnelle avec celles d'un essai randomisé et contrôlé, en mimant l'essai randomisé hypothétique que l'on souhaiterait mener pour répondre à une question causale. Cette méthodologie d'essai clinique requiert une grande rigueur dans la conception et la mise en œuvre du protocole, le recours à des méthodes d'ajustement statistique appropriées et une analyse de la robustesse de ses résultats en termes de biais de confusion résiduel, afin de lever une partie des biais inhérents aux études observationnelles.

Essai clinique simulé, également appelé in silico : essai clinique effectué à l'aide de simulations numériques pour tester un médicament, un dispositif médical ou une intervention, dans des conditions bien définies, à l'aide de modèles vérifiés et validés selon des protocoles bien définis. Le *In silico* désigne une expérimentation ou intervention par simulations numériques (dans l'ordinateur), par opposition à *in vivo* (chez l'organisme vivant, animal ou humain) ou *in vitro* (en laboratoire). L'expression s'applique aussi bien à l'échelle d'une expérience ou intervention prête à confusion biologique simple qu'à celle d'un patient, d'une étude clinique complète ou d'une cohorte de patients.

Estimand : Un estimand est une quantité à estimer dans une analyse statistique. Ce terme permet de distinguer l'objectif d'inférence de la méthode utilisée pour obtenir une approximation de cet objectif (l'estimateur) et de la valeur spécifique obtenue à partir d'une méthode et d'un ensemble de données donnés (l'estimation).

MAP - meta-analytic-predictive ou approche méta-analytique-prédictive. Cette technique tient compte, à l'aide d'un modèle hiérarchique, de l'hétérogénéité entre les essais afin de dériver un a priori informatif à partir de données historiques.

QBA : analyse des biais quantitatifs. Méthodes constituant une gamme d'approches qui utilisent des informations collectées en dehors d'un ensemble de données d'étude primaire pour mieux comprendre l'impact du biais résiduel sur les estimations de l'effet du traitement.

Modèle d'IA à usage général (IA générative) : « un modèle d'IA, y compris lorsqu'il est formé à l'aide d'une grande quantité de données en utilisant l'autosupervision à l'échelle, qui présente une grande généralité et est capable d'exécuter avec compétence un large éventail de tâches distinctes, quelle que soit la manière dont le modèle est mis sur le marché, et qui peut être intégré dans divers systèmes ou applications en aval, à l'exception des modèles d'IA qui sont utilisés pour des activités de recherche, de développement ou de prototypage avant d'être mis sur le marché » (article 3 du règlement européen sur l'IA du 13 juin 2024).

Responsable de traitement : « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement; lorsque les finalités et les moyens de ce traitement sont déterminés par le droit de l'Union ou le droit d'un État membre, le responsable du traitement peut être désigné ou les critères spécifiques applicables à sa désignation peuvent être prévus par le droit de l'Union ou par le droit d'un État membre. » (article 4 du RGPD)

Utilisateur de données : « une personne physique ou morale, y compris les institutions, organes ou organismes de l'Union, qui s'est vu octroyer un accès licite aux données de santé électroniques à des fins d'utilisation secondaire en vertu d'une

autorisation de traitement de données, d'une approbation d'une demande de données de santé ou d'une approbation d'accès émanant d'un participant autorisé à DonnéesDeSanté@UE (HealthData@EU) » (article 2 du règlement EEDS).

Utilisation secondaire : le traitement de données de santé électroniques aux fins énoncées au chapitre IV du règlement EEDS, autres que les finalités initiales pour lesquelles ces données ont été collectées ou produites (article 2 du règlement EEDS).

Système d'IA (SIA) : « un système basé sur une machine qui est conçu pour fonctionner avec différents niveaux d'autonomie et qui peut faire preuve d'adaptabilité après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des données qu'il reçoit, comment générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer des environnements physiques ou virtuels » (article 3 du règlement sur l'IA).

Sous-traitant : « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui traite des données à caractère personnel pour le compte du responsable du traitement » (article 4 du RGPD).

ACRONYMES

AIS	Agence de l'Innovation en Santé
AMM	Autorisation de mise sur le marché
ANSM	Agence Nationale du Médicament et des produits de santé
CESREES	Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé
CNIL	Commission Nationale Informatique et Liberté
CPP	Comité de protection des personnes
CT	Commission de la Transparence
DAG	Graphe orienté acyclique (en anglais, directed acyclic graph)
EC	Essai clinique
ECOG	Eastern Cooperative Oncology Group
ECR	Essai contrôlé randomisé
EHR	Electronic Health Records, également appelés dossiers patients informatisés
EMA	European Medicine Agency, l'agence du médicament Européenne
ESME	Entrepôt de Données porté par UNICANCER – « Epidémiologie-Stratégie Médico-Economique »
EUnetHTA	EUropean network for Health Technology Assessment
F-CRIN	French Clinical Research Infrastructure Network – Infrastructure nationale de recherche clinique
FDA	Food and Drug Administration – Agence de régulation des produits de santé des Etats Unis d'Amérique
GT	Groupe de Travail
HAS	Haute Autorité de Santé
IA	Intelligence artificielle
ICH	International Council for Harmonisation
MR	Méthodologie de référence
NLP	Natural Language Processing
PDS	Plateforme des données de santé (ou Health data hub)
PFS	Progression free survival ou survie sans progression
PRINCIPLED	A Process guide for INferential studies using healthcare data from routine Clinical Practice to evaluate causal Effects of Drugs
PROMs	Patient-reported outcome measures – Mesures et résultats rapportés (collectés) directement par les patients, notamment dans une recherche clinique
RECIST	Response Evaluation Criteria In Solid Tumors
RIPH	Recherche impliquant la personne humaine
RNIPH	Recherche n'impliquant pas la personne humaine
RWD	Real world data ou données de vie réelle ou de soins courants
SFPT	Société Française de Pharmacologie et de Thérapeutique
SNDS	Système National de Données de Santé

REFERENCES

- US Food and Drug Administration:
 - General glossary: <https://www.fda.gov/science-research/artificial-intelligence-and-medical-products/fda-digital-health-and-artificial-intelligence-glossary-educational-resource>
 - Additional glossary, page 26: [*Using Artificial Intelligence & Machine Learning in the Development of Drug and Biological Products*](#)
- Avicenna Alliance Glossary: <https://www.avicenna-alliance.com/glossary.html>
- Guideline ICH E10 : https://database.ich.org/sites/default/files/E10_Guideline.pdf
- LIL : Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés dite « loi Informatique et libertés »
- RGPD : règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).
- RIA ou règlement sur l'IA : règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle)

Annexe 2 Enjeux réglementaires de l'usage de méthodologies complémentaires et/ou alternatives aux essais cliniques

Sommaire

1. Introduction et contexte

2. Présentation du cadre juridique applicable aux différents types d'études

- 2.1. Les recherches organisées et pratiquées sur des personnes
 - 2.1.1. Les différents types de recherches
 - 2.1.2. Les modalités d'information et d'accord de la personne pour participer à la recherche
 - 2.1.3 La protection des données personnelles de la personne se prêtant à la recherche
 - 2.1.4. Les procédures de mise en œuvre
- 2.2. Les recherches réalisées à partir de données déjà collectées
 - 2.2.1. Les principales sources de données
 - 2.2.2. La protection des données à caractère personnel
 - 2.2.3. Les conditions de réutilisation des données
- 2.3. Les recherches réalisées à partir d'échantillons biologiques déjà collectés
 - 2.3.1. La possibilité et les conditions de réutilisation d'échantillons biologiques
 - 2.3.2. Les formalités applicables
- 2.4. Principes communs à la protection des données à caractère personnel dans les différents types de recherches
 - 2.4.1 Champ d'application de la réglementation et principes applicables à tous les types de recherches
 - 2.4.2. Synoptique des démarches réglementaires de mise en œuvre selon la catégorie de recherche
 - 2.4.3. L'anonymisation des données à caractère personnel

3. Les questions posées et les perspectives relatives aux nouvelles méthodologies d'étude

- 3.1. Les études combinant des catégories d'étude classiques
- 3.2. Les études utilisant des outils basés sur l'Intelligence artificielle

1. Introduction et contexte

L'essor du numérique et de l'intelligence artificielle permet d'envisager des schémas et des conditions et protocoles et d'essais cliniques différents du modèle de référence, le « Randomised clinical trial », qui viennent compléter l'arsenal méthodologique. Ces nouveaux schémas peuvent nécessiter la réutilisation de données de santé à caractère personnel (et/ou d'échantillons biologiques) collectés, par exemple, dans le cadre de la prise en charge médicale ou de projets de recherche. Le recours à des systèmes d'intelligence artificielle permet également de générer des données dites virtuelles ou artificielles à partir de ces données de santé.

La réglementation applicable en France prévoit des dispositions distinctes issues du code de la santé publique et de règlements européens pour :

- Les recherches organisées et pratiquées sur des personnes ;
- Les recherches nécessitant une réutilisation de données préalablement recueillies dans le cadre de la prise en charge médicale habituelle ou d'une recherche antérieure ;
- Les recherches portant sur des échantillons préalablement prélevés et collectés dans le cadre de la prise en charge médicale habituelle ou lors d'une précédente recherche.

L'application de ces réglementations (soit uniquement nationales soit également européennes) implique des interlocuteurs différents tant du côté opérationnel que pour les autorités compétentes, des procédures et des délais différents.

Ce triple encadrement juridique est moins problématique lorsque les recherches relèvent de par leur conception, d'une seule réglementation. Toutefois, certaines recherches impliquent de combiner plusieurs composantes relevant de régimes réglementaires différents (l'une impliquant une intervention sur les personnes et une autre une réutilisation de données ou

d'échantillons biologiques). Dans cette hypothèse, un besoin de clarification voire d'adaptation des réglementations apparaît nécessaire. L'objet du présent document est donc de dresser un état des lieux de la réglementation applicable.

La particularité de ces nouvelles méthodologies non courantes en recherche clinique avec adjonction de données réutilisées est d'être transverse et de mobiliser l'application de plusieurs réglementations.

2. Présentation du cadre juridique applicable aux différents types d'études

Ce chapitre présente les différents types d'études cliniques et le cadre juridique national et européen dans lequel elles s'inscrivent.

2.1. Les recherches organisées et pratiquées sur des personnes

2.1.1. Les types de recherches

L'ordonnance n° 2016-800 du 16 juin 2016 relative aux recherches impliquant la personne humaine et son décret d'application prévoient que sont considérées comme des recherches impliquant la personne humaine (« RIPH ») « les recherches organisées et pratiquées sur des personnes volontaires saines ou malades, en vue du développement des connaissances biologiques ou médicales qui visent à évaluer :

1. Les mécanismes de fonctionnement de l'organisme humain, normal ou pathologique ;
2. L'efficacité et la sécurité de la réalisation d'actes ou de l'utilisation ou de l'administration de produits dans un but de diagnostic, de traitement ou de prévention d'états pathologiques » (article R1121-1 du code de la santé publique). »

Ces recherches répondent donc à deux critères cumulatifs :

- L'un sur les modalités : la recherche est organisée et pratiquée sur des personnes humaines (volontaires saines ou malades) ;
- L'autre sur la finalité : la recherche est organisée en vue du développement des connaissances biologiques ou médicales.

Les recherches ne respectant pas ces deux critères sont considérées comme des « recherches n'impliquant pas la personne humaine » (RNIPH), terme non expressément défini dans le code de la santé publique et regroupant d'une part des recherches organisées et pratiquées sur des personnes, d'autre part des recherches réutilisant des données ou des échantillons biologiques.

Pour les recherches organisées et pratiquées sur des personnes, il existe trois catégories :

1. Les recherches impliquant la personne humaine (RIPH)

Ces recherches sont listées au sein de l'article L.1121-1 du code de la santé publique, en fonction des risques et contraintes que ces recherches induisent sur le participant :

- Les **recherches interventionnelles qui comportent une intervention sur la personne non justifiée par sa prise en charge habituelle** (dites « RIPH de catégorie 1 ») ;
- Les **recherches interventionnelles qui ne comportent que des risques et des contraintes minimales**, dont la liste est fixée par arrêté du ministre chargé de la Santé¹, après avis du directeur général de l'Agence nationale de sécurité du médicament et des produits de santé (« RIPH de catégorie 2 ») ;
- Les **recherches non interventionnelles qui ne comportent aucun risque ni contrainte dans lesquelles tous les actes sont pratiqués et les produits utilisés de manière habituelle** (« RIPH de catégorie 3 »). Certaines d'entre elles sont visées à l'article R. 1121-2 du Code de la santé publique lorsqu'elles concernent des produits mentionnés à l'article L. 5311-1 du même code. D'autres sont listées au sein d'un arrêté du ministre chargé de la santé².

Ces trois catégories de recherches sont régies par les dispositions du titre II « Recherches impliquant la personne humaine » du livre Ier de la première partie du Code de la santé publique³ (articles L.1121-1 et suivants).

2. Les recherches sur des produits de santé encadrées par des règlements européens

L'entrée en application de trois règlements européens a conduit à exclure de la catégorie des « RIPH » les recherches suivantes :

¹ Arrêté du 12 avril 2018 fixant la liste des recherches mentionnées au 2° de l'article L. 1121-1 du code de la santé publique

² Arrêté du 12 avril 2018 fixant la liste des recherches mentionnées au 3° de l'article L. 1121-1 du code de la santé publique.

³ Ordonnance n° 2016-800 du 16 juin 2016 (dernière modification le 29 juillet 2022) relative aux recherches impliquant la personne humaine

- Les **essais cliniques de médicaments** régis par les dispositions du règlement (UE) n° 536/2014 du Parlement européen et du Conseil du 16 avril 2014 ;
- Les **investigations cliniques de dispositifs médicaux** mentionnés à l'article premier du règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 régis par les dispositions de ce règlement ;
- Les **études de performances de dispositifs médicaux de diagnostic *in vitro*** mentionnés à l'article premier du règlement (UE) 2017/746 du Parlement européen et du Conseil du 5 avril 2017 régies par les dispositions de ce règlement.

Ces recherches sont donc régies par ces règlements européens et les dispositions du titre II « Recherches impliquant la personne humaine » du Livre Ier de la première partie du code de la santé publique ne s'appliquent pas à ces trois catégories de recherches.

Par exception, le titre II « Recherches impliquant la personne humaine » du livre Ier de la première partie du Code de la santé publique prévoit l'application de certaines dispositions spécifiques :

- Aux essais cliniques de médicaments (article L.1124-1 du code de la santé publique) ;
- Aux investigations cliniques de dispositifs mentionnés à l'article premier du règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 (articles L1125-1 à L1125-31 du code de la santé publique) ;
- Aux études des performances de dispositifs mentionnés à l'article premier du règlement (UE) 2017/746 du Parlement européen et du Conseil du 5 avril 2017 (articles L1126-1 à L1126-29 du code de la santé publique).

3. Les autres recherches organisées et pratiquées sur des personnes ne répondant pas aux finalités des RIPH

La finalité de ces recherches n'est pas le développement des connaissances biologiques ou médicales tel que précisé dans l'article R. 1121-1 du Code de la santé publique. Il en est ainsi par exemple des enquêtes de satisfaction du consommateur pour des produits cosmétiques ou alimentaires, des expérimentations en sciences humaines et sociales dans le domaine de la santé. Ces recherches ne sont pas régies par des dispositions spécifiques issues du code de la santé publique ou d'un des trois règlements européens susmentionnés.

2.1.2. Les modalités d'information et d'accord de la personne pour participer à la recherche

S'agissant des recherches impliquant la personne humaine, l'article L.1121-1-3 du code de la santé publique impose, par principe la délivrance d'une information individuelle ainsi que :

- Le recueil d'un consentement (libre, éclairé et selon les cas, écrit ou exprès) en vue de la participation de l'étude pour les RIPH de catégories 1⁴ et 2 ;
- L'absence d'opposition pour les RIPH de catégorie 3.

Ce consentement ou cette absence d'opposition doit émaner des personnes se prêtant à la recherche et/ou les personnes, organes ou autorités chargés de les assister, de les représenter ou d'autoriser la recherche au sens de l'article L. 1122-1 du code de la santé publique (membre de la famille ou personne de confiance en cas d'inclusions en urgence ; titulaires de l'exercice de l'autorité parentale pour les mineurs, etc.).

Il est à noter par ailleurs que les dispositions de l'article L.1122-1-2 du Code de la santé publique prévoient que le promoteur d'une RIPH peut demander à la personne se prêtant à une recherche d'accepter « par principe » que ses données soient utilisées pour des projets de recherches ultérieurs. Dans cette hypothèse, le participant peut retirer son consentement à tout moment et doit être informé de chaque projet de recherche ultérieur afin de pouvoir, le cas échéant, exercer son droit d'opposition. Un mécanisme similaire est également prévu par le règlement européen (UE) 536/2014 relatif aux essais cliniques de médicaments à usage humain.

Les dispositions de l'article 29 du règlement 536/2014 prévoient que **les essais cliniques de médicaments** ne peuvent être mis en œuvre qu'après le recueil du consentement éclairé des participants (ou de leur représentant légal). Ce principe souffre d'une exception prévue à l'article 30 du règlement : ce consentement peut être recueilli de façon simplifiée pour les essais par grappes ou « cluster randomization trial », l'essai clinique étant conduit en randomisant des groupes de sujets plutôt que des sujets eux-mêmes.

Les **investigations cliniques** portant sur les dispositifs médicaux ne peuvent également être mises en œuvre qu'après le recueil du consentement éclairé du participant ou de leur représentant légal (article 63 du règlement 2017/745), tout comme les **études des performances sur les dispositifs médicaux de diagnostic *in vitro*** (article 59 du règlement (UE) 2017/746).

⁴ Pour aider les promoteurs, une trame de [note d'information](#) a été élaborée par la CNRIPH en lien avec la DGS et la CNIL.

Certaines **dispositions spécifiques** aménagent, dans certains cas, la possibilité de limiter les effets associés au retrait du consentement sur l'utilisation des données. Il en va ainsi par exemple des dispositions de l'article L.1122-1-1 du code de la santé publique applicables aux RIPH qui prévoient que « dans le cas où la personne se prêtant à une recherche a retiré son consentement, ce retrait n'a pas d'incidence sur les activités menées et sur l'utilisation des données obtenues sur la base du consentement éclairé exprimé avant que celui-ci n'ait été retiré. » Le considérant 76 du règlement sur les essais cliniques de médicaments est convergent sur ce sujet.

Les personnes concernées doivent en être informées, afin que leur consentement pour participer à l'étude soit valablement recueilli.

2.1.3. La protection des données personnelles de la personne se prêtant à la recherche

L'application du RGPD et de la Loi « Informatique et Libertés »

Le RGPD et la loi « Informatique et Libertés » s'appliquent aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches quelle que soit leur méthodologie, sous réserve qu'ils relèvent du champ d'application matériel et territorial du règlement et de la loi. Il appartient alors au responsable de traitement de documenter la conformité de son projet de recherche en santé vis-à-vis du RGPD et de la loi « Informatique et libertés » à l'aide des questions présentées dans le paragraphe 2.4.

Les personnes dont les données à caractère personnel ont été recueillies doivent être individuellement informées du traitement de leurs données conformément aux dispositions du RGPD et à l'article 69 de la loi « Informatique et libertés ».

Il est à noter que le RGPD impose une information des personnes concernées par ces projets de recherche concise, transparente, compréhensible et aisément accessible en des termes clairs et simples. Ces projets de recherche impliquant une collecte de données à caractère personnel auprès des personnes concernées, il convient de faire application de l'article 13 du RGPD qui prévoit les mentions devant figurer dans la note d'information (identité du responsable de traitement, finalités du traitement, destinataires des données, durée de conservation des données *etc*⁵).

Concernant les paragraphes 2.1.2 et 2.1.3 et pour aider les promoteurs, un outil d'aide à la rédaction d'un document d'information à l'attention des participants à une recherche impliquant la personne humaine est disponible sur le site du ministère de la santé⁶.

2.1.4. Les procédures de mise en œuvre

Les **recherches impliquant la personne humaine** ne peuvent être mises en œuvre qu'après l'avis favorable d'un comité de protection des personnes (CPP) ainsi que, le cas échéant, l'autorisation de l'Agence nationale de sécurité du médicament et des produits de santé (ANSM) notamment pour les recherches de « catégorie 1 » et les recherches interventionnelles portant sur un produit de santé.

Sont également soumis à un examen scientifique et éthique :

- Les **essais cliniques de médicaments** en vue de l'obtention d'une autorisation préalable par les organismes compétents conformément à l'article 4 du règlement (UE) 536/2014 du 16 avril 2014⁷ ;
- Les **investigations cliniques** liées aux dispositifs médicaux conformément à l'article 62.3 du règlement (UE) 2017/745 du 5 avril 2017⁸ ;
- Les **études des performances de dispositifs médicaux de diagnostics *in vitro*** conformément à l'article 58.3 du règlement (UE) 2017/746 du 5 avril 2017⁹.

Par ailleurs, conformément aux dispositions de l'article 66 de la loi 78-17 du 6 janvier 1978 modifiée « Informatique et libertés », les traitements de données mis en œuvre dans le cadre des recherches, études ou évaluations dans le domaine de la santé (RIPH, essais cliniques, investigations cliniques, études des performances) doivent faire l'objet d'une déclaration de conformité à une méthodologie standard de traitement de données décrite dans un référentiel publié par la CNIL appelé « méthodologie de référence ».

Par exception, les recherches qui ne sont pas conformes à ces référentiels en raison d'une sensibilité particulière (traitements de données d'infractions par exemple) doivent être autorisées par la CNIL, après avis du comité de protection des personnes¹⁰.

⁵ Voir la fiche pratique « Quelles formalités pour les traitements de données de santé ? publiée par la CNIL.

⁶ Outil d'aide à la rédaction d'un document d'information élaboré par la CNRIPH en lien avec la DGS et la CNIL.

⁷ Voir la fiche pratique « Médicaments – Demander une autorisation pour un essai clinique » publiée par l'ANSM.

⁸ Voir la fiche pratique « Dispositifs médicaux - Demander une autorisation pour une investigation clinique » publiée par l'ANSM.

⁹ Voir la fiche pratique « Dispositifs médicaux de diagnostic *in vitro* – Demander une autorisation pour une étude des performances » publiée par l'ANSM.

¹⁰ Sur ce point, voir la fiche pratique « Recherches dans le cadre de la santé : quelles sont les formalités ? » publiée par la CNIL.

Si la recherche est qualifiée de recherche impliquant la personne humaine (RIPH)

MR-001
adoptée le 3 mai 2018

Applicable si la recherche nécessite le recueil du consentement des personnes concernées pour réaliser la recherche.

Cela correspond :

- aux recherches interventionnelles sur la personne non justifiée par sa prise en charge habituelle (RIPH1) ;
- aux recherches interventionnelles qui ne comportent que des risques et contraintes minimales (RIPH2) pour lesquelles une information individuelle est réalisée ;
- aux recherches impliquant le recueil du consentement en raison de la réalisation d'un examen des caractéristiques génétiques des participants ;
- aux essais cliniques de médicaments (régis par le règlement (UE) 536/2014 du 16 avril 2014 relatif aux essais cliniques de médicaments à usage humain), à l'exception des essais cliniques par grappes ;
- aux investigations cliniques liées aux dispositifs médicaux (régies par le règlement (UE) 2017/745 du 5 avril 2017 relatif aux dispositifs médicaux).
- aux tests de performance de diagnostics in vitro (régis par le règlement (UE) 2017/746 du 5 avril 2017 relatif aux dispositifs médicaux de diagnostic in vitro).

MR-003
adoptée le 3 mai 2018

Applicable aux recherches nécessitant une information des personnes concernées sans que ces dernières ne s'y opposent. Cela correspond :

- aux recherches interventionnelles qui ne comportent que des risques et contraintes minimales (RIPH2) pour lesquelles une information collective est réalisée, après avis du comité de protection des personnes ;
- aux recherches non interventionnelles qui ne comportent aucun risque ni contrainte sur la personne (RIPH3) ;
- aux essais de médicaments par grappes.

Figure 1 : Liste des référentiels applicables aux recherches conduites sur la personne hormis celles ne répondant pas aux finalités des RIPH.

Le schéma général d'articulation des procédures réglementaires à mettre en œuvre en fonction de la qualification de la recherche est présenté en section 2.4.2.

2.2. Les recherches réalisées à partir de données déjà collectées

Ces recherches nécessitent de réutiliser des données déjà collectées dans le cadre de la prise en charge habituelle ou d'une précédente recherche (données issues des dossiers médicaux, d'un registre, d'un entrepôt de données de santé, etc.).

Ces recherches, qui ne respectent ni les modalités ni les finalités prévues pour les « recherches impliquant la personne humaine », entrent dans la catégorie des « recherches n'impliquant pas la personne humaine (RNIPH) ».

2.2.1. Les principales sources de données

Les dossiers médicaux

Les données de santé des patients collectées par un établissement de santé, public ou privé à l'occasion d'une prise en charge sanitaire, médico-sociale ou des activités nécessaires à la coordination de ces prises en charge sont enregistrées et conservées de manière nominative au sein du dossier médical. Lorsque cette conservation se fait de façon dématérialisée, il s'agit du dossier patient informatisé ((DPI), en anglais « *electronic medical records EMRs* ». Cette collecte de données par les professionnels de santé dans le cadre de la prise en charge des patients n'est pas soumise à la réalisation d'une formalité préalable auprès de la CNIL. Les données contenues dans les dossiers médicaux peuvent être réutilisées dans le cadre de projets de recherche.

Les entrepôts de données de santé

Les entrepôts de données de santé sont des bases de données destinées à être utilisées notamment à des fins de recherches, d'études ou d'évaluations dans le domaine de la santé.

Lorsqu'un organisme envisage de constituer une importante base de données centralisée (dénommée « entrepôt de données de santé ») afin de réaliser ultérieurement plusieurs projets de recherches, un raisonnement en deux temps doit être opéré. En effet, chaque traitement fait l'objet d'un régime juridique distinct :

- La création de l'entrepôt de données en tant que tel (c'est-à-dire la collecte et la conservation des données dans une base unique pendant une plus longue durée) ;
- Les projets de recherches, d'études ou d'évaluations réalisés à partir des données conservées dans l'entrepôt par le même responsable de traitement ou d'autres organismes.

La CNIL a publié une fiche pratique pour aider les organismes à qualifier le traitement qu'ils envisagent de mettre en œuvre¹¹.

La notion « d'entrepôt »

Les entrepôts de données de santé sont créés principalement **pour collecter et disposer de données massives**

- **Origine variée des données** (données relatives à la prise en charge médicale du patient, données socio-démographiques, données issues de précédentes recherches etc.)
- **Longue prolongée** de l'entrepôt
- **Base de données alimentée au fil de l'eau**
- Données réutilisées à la réalisation de **traitements ultérieurs**

Entrepôt	Recherche
Permet la réalisation ultérieure d'un nombre important de projets (dont les finalités sont diverses)	Finalité précise et répond à une question de recherche scientifique spécifique et ponctuelle
Constitué afin d'obtenir un volume de données important.	Les données sont collectées spécifiquement pour les besoins du projet.
Constitué pour une durée assez longue (10 ans en général)	La durée de la recherche est limitée et connue

Figure 2, Distinction entre entrepôt et recherche sur les données, Source : CNIL

S'agissant de la constitution de l'entrepôt, trois hypothèses doivent être envisagées :

- Le **consentement explicite** des personnes concernées par la collecte, l'enregistrement et la conservation des données de santé dans un entrepôt a été recueilli. Dans ce cas, aucune formalité n'est nécessaire mais le responsable de traitement devra tout de même être en mesure de démontrer la conformité de son traitement au RGPD (registre, analyse d'impact, etc.) ;
- Le responsable de traitement réalise un **engagement de conformité au référentiel sur les entrepôts de données de santé**¹². La conformité au référentiel ne peut être invoquée que si l'entrepôt permet au responsable de traitement d'exercer une mission d'intérêt public (base légale du traitement) et si le traitement envisagé est conforme en tous points au référentiel ;
- **Si le responsable de traitement ne remplit aucune de deux conditions présentées ci-dessus**, le traitement relatif à la constitution de l'entrepôt doit faire l'objet d'une demande d'autorisation « santé » (hors recherche) auprès de la CNIL.

Plusieurs dizaines d'entrepôts ont ainsi été constitués en France et sont recensés dans la cartographie publiée par la CNIL¹³.

¹¹ Voir la fiche pratique « Traitements de données de santé : comment faire la distinction entre un entrepôt et une recherche et quelles conséquences » publiée par la CNIL.

¹² Sur ce point, voir la check-list de conformité au référentiel « entrepôt de données de santé » publiée par la CNIL.

¹³ Sur ce point, voir la fiche pratique « Explorez la cartographie des entrepôts de données de santé en France » publiée par la CNIL.

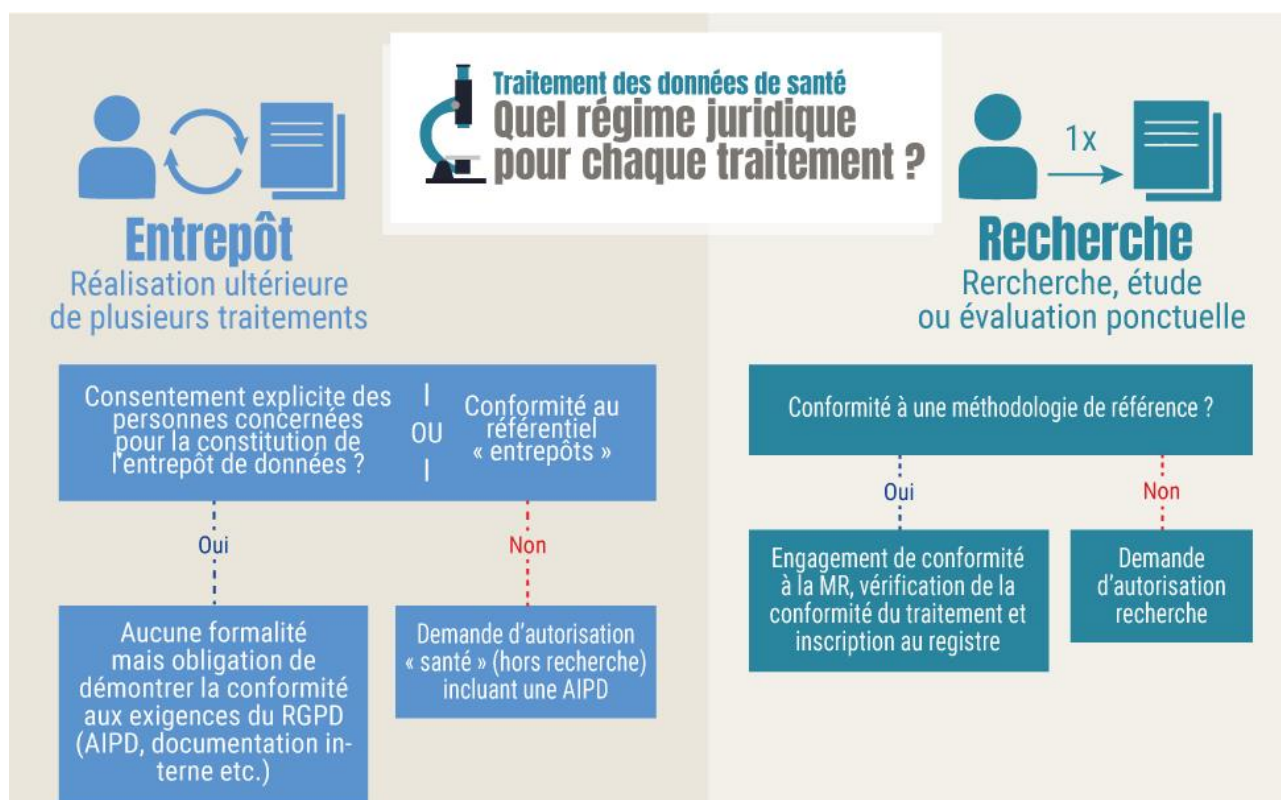


Figure 3, Régimes juridiques liés aux entrepôts et aux recherches sur les données ; Source : CNIL

Depuis la loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé, la CNIL peut autoriser la constitution d'entrepôts contenant des données du système national des données de santé (seules ou associées avec des données issues d'autres traitements). Dans cette hypothèse, une demande d'autorisation doit systématiquement être déposée auprès de la CNIL lorsque l'entrepôt comporte des données du SNDS, même lorsque le consentement des personnes concernées a été recueilli¹⁴.

Le Système national des données de santé

Le SNDS a été créé par la loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé afin de développer l'usage des données de santé, notamment à des fins de recherches dans le domaine de la santé. Historiquement, il est notamment constitué :

- Des données de remboursement de l'Assurance maladie (système national d'information interrégimes de l'Assurance Maladie, le SNIIRAM) ;
- Des données médico-administratives hospitalières (base du programme de médicalisation des systèmes d'information – PMSI) ;
- Des données sur les causes médicales de décès (base du centre d'épidémiologie sur les causes médicales de décès de l'institut national de santé et de la recherche médicale – CépiDc de l'Inserm) ;
- Des données relatives au handicap (en provenance des maisons départementales des personnes handicapées – MDPH - données de la caisse nationale de solidarité pour l'autonomie - CNSA). Ces données font partie du « SNDS historique ».

La loi n° 2019-774 du 24 juillet 2019 a élargi le périmètre du SNDS à des catégories de données complémentaires comme les données des enquêtes dans le domaine de la santé, les données de protection maternelle et infantile, etc.

L'ensemble des données visées à l'article L.1461-1 du Code de la santé publique entre dans le périmètre du « SNDS élargi ». Lorsqu'elles sont utilisées pour l'une des finalités visées à l'article L. 1461-1 III du Code de la santé publique, elles sont soumises au respect de l'ensemble des dispositions de ce code, tout particulièrement au respect du référentiel de sécurité¹⁵, à l'interdiction de poursuite de finalités interdites visées au V de l'article L.1461-1, aux conditions d'accès aux données via un accès permanent ou à la réalisation d'une formalité auprès de la CNIL ainsi qu'aux obligations de transparence.

¹⁴ Sur ce point, voir la [fiche pratique](#) « Demande d'autorisation dans le domaine de la santé (hors recherche) : les informations à fournir et les critères d'octroi » publiée par la CNIL.

¹⁵ Sur ce point, voir l'[arrêté du 6 mai 2024](#) relatif au référentiel de sécurité applicable au SNDS.

L'essentiel de ces données a vocation à être intégré progressivement dans un « SNDS centralisé » par la Caisse nationale de l'assurance maladie (CNAM) et la Plateforme des données de santé (PDS ou *Health data hub* – HDH). Ce SNDS centralisé est composé :

- D'une **base principale** couvrant l'ensemble de la population (comprenant à ce jour le « SNDS historique » et des données relatives à la Covid-19 extraites des bases Vaccin Covid et SI-DEP – le système d'information et de dépistage) ;
- D'une **base catalogue** incluant d'autres bases de données considérées comme pertinentes pour les acteurs de la recherche. A ce jour, une vingtaine de bases ont été inscrites au catalogue du SNDS. Parmi ces bases, figure notamment la base « EpiCov », enquête statistique sur l'état de santé et les conditions de vie en lien avec l'épidémie due au coronavirus, la base « HEPATHER » relative aux patients atteints d'hépatite B ou C, la base « OSCOUR » sur les données de passage aux urgences ou encore la base de données « BNDMR » relative à la banque nationale des maladies rares centralisant les dossiers patients informatisés créés par les centres experts¹⁶.

Les données de la base principale ont vocation à être stockées et mises à disposition par la CNAM et la Plateforme des données de santé (PDS). Cette dernière est également responsable du stockage et de la mise à disposition des données du catalogue du SNDS.

Ces données (seules ou appariées de façon déterministe/directe¹⁷ ou probabiliste/indirecte avec d'autres données issues du soin ou de précédentes études) peuvent être réutilisées à des fins de recherches, d'études ou d'évaluation dans le domaine de la santé ou alimenter des entrepôts de données de santé.

Les registres

Au sens épidémiologique, un registre de santé est historiquement défini comme un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs événements de santé dans une population géographiquement définie, à des fins de recherche de santé publique par une équipe ayant les compétences appropriées. Cette définition était prévue par un arrêté du 6 novembre 1995 abrogé en 2013.

De tels registres ont trois rôles : l'amélioration des connaissances scientifiques ; l'évaluation des actions entreprises pour lutter contre les maladies ; l'estimation de la prévalence. Depuis 2013, les registres sont répertoriés par le Comité d'évaluation des registres (CER), qui associe l'Agence nationale de santé publique, l'Institut national de la recherche médicale (Inserm) et l'Institut national du cancer (INCa). Le CER a pour mission de coordonner l'action des différents registres, de les évaluer et de donner un avis sur toute demande de qualification d'un nouveau registre.

D'un point de vue « informatique et libertés », les registres de pathologies constituent la plupart du temps des entrepôts de données de santé soumis au cadre juridique mentionné précédemment. Dans ce cadre, la CNIL accompagne la mise en conformité juridique et technique de plusieurs registres.

2.2.2. La protection des données à caractère personnel

Le RGPD et la loi « Informatique et Libertés » s'appliquent aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches quelle que soit leur méthodologie, sous réserve qu'ils relèvent du champ d'application matériel et territorial du règlement et de la loi. Il appartient alors au responsable de traitement de documenter la conformité de son projet de recherche en santé vis-à-vis du RGPD et de la loi « Informatique et libertés » à l'aide des questions présentées dans le paragraphe 2.4.

Par ailleurs, le règlement (UE) 2025/327 du Parlement européen et du Conseil du 11 février 2025 relatif à l'espace européen des données de santé (dit règlement EEDS)¹⁸ est entré en vigueur le 26 mars 2025. Son chapitre IV, dont les principales dispositions seront applicables à compter du 26 mars 2029, ont vocation à régir l'utilisation secondaire de certaines catégories données de santé électroniques (présentant ou non un caractère personnel) visées à l'article 51.

Le chapitre IV du règlement EEDS impose aux détenteurs d'un grand nombre de catégories de données de santé (à l'exception des personnes physiques, y compris les chercheurs individuels et des micro-entreprises au sens de la recommandation 2003/361/CE en application de l'article 50), de **mettre à disposition les données** pour certaines finalités strictement énumérées à l'article 53 du règlement EEDS¹⁹.

¹⁶ Sur ce point, voir l'[arrêté du 12 mai 2022](#) relatif aux données alimentant la base principale et aux bases de données du catalogue du système national des données de santé et son [arrêté modificatif](#) en date du [7 février 2025](#).

¹⁷ Sur ce point, voir les [fiches pratiques](#) « Les circuits d'appariement avec le SNDS utilisant le NIR » publiées par la CNIL.

¹⁸ [Règlement \(UE\) 2025/327 du Parlement européen et du Conseil du 11 février 2025 relatif à l'espace européen des données de santé et modifiant la directive 2011/24/UE et le règlement \(UE\) 2024/2847 \(Texte présentant de l'intérêt pour l'EEE\)](#)

Sur ce point, voir la [fiche pratique](#) « Espace européen des données de santé (EEDS) publiée par le ministère de la santé et de la prévention ».

¹⁹ Il est à noter que les dispositions de l'article 54 du règlement EEDS proscrivent la poursuite de certaines finalités dans le cadre d'une utilisation secondaire des données de santé. Il en va ainsi notamment de l'exercice d'activités de publicité ou de marketing.

Parmi les finalités prévues figure notamment « la recherche scientifique ayant trait aux secteurs de la santé ou des soins qui contribue à la santé publique ou aux évaluations des technologies de la santé, ou qui garantit un niveau élevé de qualité et de sécurité des soins de santé, des médicaments ou des dispositifs médicaux, dans le but d'en faire bénéficier les utilisateurs finals, tels que les patients, les professionnels de la santé et les administrateurs des services de santé, y compris [...] les essais et l'évaluation d'algorithmes, notamment dans les dispositifs médicaux, les dispositifs médicaux de diagnostic in vitro, les systèmes d'IA et les applications de santé numérique. »

Cette mise à disposition pourra être effectuée au format statistique anonymisé via une « demande de données de santé » (article 69 du règlement EEDS) ou impliquer un accès à des données à caractère personnel dans le cadre d'une « demande d'accès aux données de santé » prévue à l'article 67 du règlement EEDS.

Si l'article 60 du règlement EEDS impose un certain nombre d'obligations aux détenteurs de données de santé en matière de mise à disposition des données et d'information sur les données disponibles, certaines garanties sont également prévues :

- La perception de redevances pour la mise à disposition des données dans les conditions prévues à l'article 62 du règlement EEDS, étant entendu que cette redevance peut inclure une compensation pour les coûts exposés par le détenteur de données de santé pour la compilation et la préparation des données de santé électroniques mises à disposition ;
- La protection des droits de propriété intellectuelle et des secrets des affaires dans les conditions prévues à l'article 52 du règlement EEDS.

D'un point de vue procédural, le règlement prévoit aussi la mise en place de guichets uniques nationaux via la désignation d'organismes responsables de l'accès aux données (ORAD) dont les missions sont listées à l'article 57 du règlement EEDS. Ils sont notamment chargés de :

- Délivrer les autorisations d'accès aux données à des fins d'utilisation secondaire dans les conditions prévues à l'article 68 du règlement EEDS ;
- De mettre les données à disposition des utilisateurs dans des conditions sécurisées sous forme anonymisée ou pseudonymisée et dans les délais prévus par le règlement (quatorze mois maximum pour les données personnelles et un an pour les données anonymisées) ;
- Mettre à la disposition du public certaines informations sur les conditions dans lesquelles les données de santé électroniques sont mises à disposition à des fins d'utilisation secondaire conformément aux dispositions de l'article 58 du règlement EEDS ;
- Surveiller et contrôler le respect des exigences énoncées dans le règlement EEDS par les utilisateurs de données de santé et les détenteurs de données de santé.

Outre l'accroissement de la transparence sur les traitements mis en œuvre, le règlement renforce les exigences de transparence sur **les données disponibles** (via des catalogues de métadonnées mis à disposition par l'ORAD) et leur qualité (via la possibilité de prévoir un label de qualité et d'utilité des données mentionné à l'article 78 du règlement EEDS).

Si l'article 105 du règlement EEDS prévoit que son entrée en application est échelonnée dans le temps, il est d'application directe. Toutefois, certaines de ses dispositions introduisent des nouveautés qui nécessiteront une adaptation du cadre législatif et réglementaire français, notamment au regard de la possibilité pour les Etats membres de maintenir des procédures nationales d'accès aux données (article 1^{er} paragraphes 7 et 8 du règlement EEDS). La Délégation au numérique en santé a lancé des travaux en ce sens dès mars 2025²⁰.

2.2.3. Les conditions de réutilisation des données

Ce paragraphe présente les principes et les conditions réglementaires de leur réutilisation, ainsi que les difficultés (lire aussi le rapport du GT AIS/F-CRIN « Pour l'utilisation des données de santé en vie réelle en recherche clinique »²¹).

- **La vérification de la licéité de la collecte initiale**

En premier lieu, il convient de vérifier que **la base de données ayant vocation à être réutilisée a été constituée et est mise en œuvre dans le respect de la réglementation applicable** (documentation de la conformité aux principes « Informatiques et libertés » décrits précédemment et aux éventuelles dispositions spécifiques issues notamment du code de la santé publique, réalisation, le cas échéant d'une formalité auprès de la CNIL²², etc.).

²⁰ Sur ce point, voir les informations publiées sur le [site web](#) du Ministère de la santé.

²¹ <https://www.fcrcin.org/nouvelles-methodologies-en-recherche-clinique/livrables-et-avancement-des-travaux>

²² Sur ce point, voir la [fiche pratique](#) « Quelles formalités pour les traitements de données de santé ? » publiée par la CNIL.

Ainsi, par exemple, si les données proviennent d'un **entrepôt de données de santé**²³, il convient de s'assurer qu'il a été constitué dans le respect du cadre réglementaire précédemment décrit (recueil du consentement, déclaration de conformité au référentiel « entrepôt de données de santé » ou autorisation de la CNIL).

- **Les modalités d'information et d'exercice des droits des personnes**

Les personnes dont les données à caractère personnel ont été recueillies doivent être **individuellement** informées du traitement de leurs données conformément aux dispositions du RGPD et à l'article 69 de la loi « Informatique et libertés ».

²⁴

Le RGPD impose une information des personnes concernées concise, transparente, compréhensible et aisément accessible en des termes clairs et simples. Les articles 13 (collecte directe) et 14 du RGPD (collecte indirecte ou lorsque la recherche implique la réutilisation de données d'une base existante) prévoient les modalités d'information et les éléments devant figurer dans la note d'information (identité du responsable de traitement, finalités du traitement, destinataires des données, durée de conservation des données *etc.*²⁵).

La méthodologie de référence MR-004 de la CNIL prévoit que des données issues du soin ou d'une précédente étude peuvent faire l'objet d'une réutilisation sans qu'il soit procédé à une nouvelle information individuelle des personnes lorsque l'information délivrée lors de la collecte initiale des données prévoit la possibilité de réutiliser les données et renvoie à un dispositif spécifique d'information auquel les personnes concernées pourront se reporter préalablement à la mise en œuvre de chaque nouveau traitement de données (par exemple : un site web ou une page dédiée appelé « portail de transparence » sur lequel serait présenté chaque projet de recherche mené sur les données collectées dans le cadre de l'information initiale). Cette méthode est vivement recommandée par la CNIL car elle constitue un moyen simple d'informer les personnes concernées sur l'ensemble des projets de recherche ultérieurs. Elle nécessite néanmoins d'avoir anticipé la possibilité de réutilisation dès la conception du traitement initial en créant une page web dédiée à l'information.

En dehors du cas où la personne concernée dispose déjà de l'intégralité des informations qui lui sont dues, d'autres exceptions au principe d'information individuelle sont prévues par le RGPD et la loi « Informatique et libertés ».

PREMIEREMENT, les dispositions de l'article 69 de la loi « Informatique et libertés » prévoient que la personne concernée ne doit pas être informée du traitement de ses données si elle a entendu faire usage du droit qui lui est reconnu par l'article L. 1111-2 du code de la santé publique d'être laissée dans l'ignorance d'un diagnostic ou d'un pronostic.

DEUXIEMEMENT, dans l'hypothèse où une nouvelle information individuelle ne pourrait être délivrée et que la réutilisation des données n'aurait pas été anticipée au moyen d'un portail de transparence, trois exceptions à la fourniture d'une information individuelle sont prévues par l'article 14.5.b) du RGPD pour les traitements mis en œuvre à des fins de recherche scientifique. Il en va ainsi lorsque la fourniture d'une information individuelle :

1. Se révèle impossible ;
2. Exigerait des efforts disproportionnés de la part du responsable de traitement au regard de l'ancienneté des données ou du nombre de personnes concernées ;
3. Est susceptible de compromettre ou de rendre impossible la réalisation des objectifs du traitement.

Des mesures appropriées doivent alors être mises en œuvre par le responsable de traitement afin de protéger les droits et libertés des personnes concernées. Dans ce cas, le RGPD, éclairé par les lignes directrices adoptées par le Comité européen de la protection des données²⁶, prévoit que des mesures appropriées doivent être prises pour protéger les droits et libertés des personnes concernées, notamment par la mise en place systématique d'une information collective via des canaux de communication appropriés au regard du contexte de l'étude (a minima, sur le site web du responsable de traitement).

TROISIEMEMENT, les dispositions de l'article 86 de la loi « informatique et libertés » prévoient que les informations concernant les personnes décédées, y compris celles qui figurent sur les certificats des causes de décès, peuvent faire l'objet d'un traitement à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé, sauf si l'intéressé a, de son vivant, exprimé son refus par écrit.

- **Les formalités applicables**

²³ La CNIL a publié une [fiche pratique](#) pour aider les responsables de traitement à qualifier leur traitement. Cette fiche pratique peut notamment servir à qualifier les registres et les cohortes.

²⁴ Pour en savoir plus, voir la [fiche pratique](#) « Traitement de données de santé : comment informer les personnes concernées ? » publiée par la CNIL.

²⁵ Par ailleurs, en présence d'une prise de décision automatisée, les personnes doivent être informées, lors de la collecte de leurs données et à tout moment sur leur demande, de l'existence d'une telle décision, de la logique sous-jacente ainsi que de l'importance et des conséquences prévues de cette décision.

²⁶ Sur ce point, voir [les lignes directrices](#) du CEPD sur la transparence au sens du règlement (UE) 2016/679.

Les traitements mis en œuvre à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé doivent faire l'objet de formalités préalables auprès de la CNIL, à l'exception des études internes²⁷.

Plusieurs référentiels ont été élaborés par la CNIL afin de faciliter la mise en œuvre des recherches, études et évaluations dans le domaine de la santé.

Liste des référentiels applicables aux recherches n'impliquant pas la personne humaine (RNIPH)	
MR-004 adoptée le 3 mai 2018	Applicable aux recherches, études ou évaluations n'impliquant pas la personne humaine (autres que les traitements de données du SNDS exclusivement).
MR-005 adoptée le 7 juin 2018	Applicable aux recherches, études ou évaluations nécessitant l'accès aux données du PMSI et/ou des RPU par les établissements de santé et les fédérations hospitalières.
MR-006 adoptée le 7 juin 2018	Applicable aux recherches, études ou évaluations nécessitant l'accès aux données du PMSI par les industriels de santé.
MR-007 adoptée le 20 juillet 2023	Applicable aux recherches, études ou évaluations nécessitant l'accès à la base principale du SNDS par les organismes agissant dans le cadre de leur mission d'intérêt public. Un <u>kit pédagogique</u> dédié est disponible sur le site internet de la Plateforme des données de santé.
MR-008 adoptée le 20 juillet 2023	Applicable aux recherches, études ou évaluations nécessitant l'accès à la base principale du SNDS par les organismes agissant dans le cadre de leurs intérêts légitimes (par exemple : industriels du secteur de la santé traitant les données du SNIIRAM). Un <u>kit pédagogique</u> dédié est disponible sur le site internet de la Plateforme des données de santé.
Référentiel ESND (ex-EGB) Adopté le 13 octobre 2022	Encadre l'accès à l'échantillon des données du SNDS (ENSD).

Ainsi, s'agissant plus spécifiquement des « recherches n'impliquant pas la personne humaine », quatre hypothèses doivent être distinguées :

1. Si la recherche est strictement conforme à la méthodologie de référence MR-004, MR-005 ou MR-006 seule une déclaration de conformité auprès de la CNIL est nécessaire ;
2. Si la recherche est strictement conforme à la méthodologie de référence MR-007 et MR-008, une déclaration de conformité auprès de la CNIL est nécessaire et un avis favorable du **comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES)** doit être recueilli²⁸ ;
3. Si la recherche est strictement conforme au référentiel portant sur la description et les garanties de procédure permettant la mise à disposition en vue de leur traitement de l'échantillon du système national des données de santé (ESND) et des bases de données thématiques appelées « datamarts » du système national d'information interrégimes de l'assurance maladie (SNIIRAM), la demande doit être approuvée par la Plateforme des données de santé (PDS) dans un délai de quinze jours ;
4. Si la recherche n'est pas conforme à un des référentiels précités, le responsable de traitement doit déposer une demande d'autorisation « recherche » en détaillant les points de non-conformité au référentiel correspondant. La demande devra être déposées auprès de la PDS, qui se chargera de la transmettre **au CESREES pour avis, puis à la CNIL pour autorisation**²⁹.

Le schéma général d'articulation des procédures réglementaires à mettre en œuvre est présenté en section 2.4.2.

2.3. Les recherches sur les échantillons biologiques

²⁷ Sur ce point, voir la fiche pratique « Recherches dans le cadre de la santé : quelles sont les formalités ? » publiée par la CNIL.

²⁸ Sur ce point, voir les guides pédagogiques élaborés par la Plateforme des données de santé et la CNIL.

²⁹ Sur ce point, voir la fiche pratique « Demande d'autorisation d'une recherche en santé : les informations à fournir et les critères d'octroi » publiée par la CNIL.

2.3.1. La possibilité et les conditions de réutilisation d'échantillons biologiques

- **Hors contexte de l'examen des caractéristiques génétiques**

Les dispositions du Code de la santé publique encadrent la possible réutilisation d'échantillons à une fin autre que celle pour lesquels ils ont été prélevés.

Elles prévoient la possible réutilisation d'éléments et de produits du corps humain à une fin médicale ou scientifique (article L.1211-2 du Code de la santé publique) sauf opposition exprimée par la personne concernée (ou son représentant légal), dûment informée au préalable de cette autre fin.

Par exception, il peut être dérogé à cette obligation d'information lorsque celle-ci se heurte à l'impossibilité de retrouver la personne concernée, ou lorsqu'un Comité de protection des personnes, consulté par le responsable de la recherche, n'estime pas cette information nécessaire. Toutefois, ces dérogations ne sont pas admises lorsque les éléments initialement prélevés consistent en des tissus ou cellules germinaux.

- **A des fins de réalisation d'un examen des caractéristiques génétiques**

Les dispositions du code de la santé publique distinguent :

- D'une part, l'examen des caractéristiques génétiques constitutionnelles définies comme l'analyse des caractéristiques génétiques d'une personne héritées ou acquises à un stade précoce du développement prénatal par l'article L.1130-1 du Code de la santé publique ;
- D'autre part, l'examen des caractéristiques génétiques somatiques définies comme la recherche et l'analyse des caractéristiques génétiques dont le caractère hérité ou transmissible est en première intention inconnu par l'article L.1130-2 du Code de la santé publique.

L'article L.1130-5 du Code de la santé publique prévoit que l'examen des caractéristiques génétiques d'une personne à des fins de recherche scientifique peut être réalisé à partir d'éléments du corps de cette personne prélevés à d'autres fins. L'article conditionne cette possibilité au fait que cette personne, dûment informée du programme de recherche, au sens de l'article L. 1243-3 du code de la santé publique, n'a pas exprimé son opposition. Cette opposition à l'examen peut être exprimée « sans forme » tant qu'il n'y a pas eu d'intervention sur l'élément concerné dans le cadre de la recherche. L'article prévoit également des modalités d'information spécifiques lorsque la personne est un mineur, fait l'objet d'une mesure de protection juridique avec représentation relative à la personne.

Enfin, lorsque la personne ne peut être retrouvée ou qu'elle est décédée ou qu'elle est hors d'état d'exprimer sa volonté, la recherche est soumise à l'avis d'un Comité de protection des personnes saisi par le responsable du programme de recherche. Le Comité est alors chargé d'évaluer les éléments justifiant de l'impossibilité de procéder à l'information de la personne et se prononce sur l'opportunité de l'examen de ses caractéristiques génétiques au regard de cette situation ainsi que de la pertinence éthique et scientifique de la recherche.

- **L'organisation d'une collection d'échantillons biologique préalable à la réutilisation**

En vue d'être utilisés à des fins scientifiques, ces échantillons biologiques humains (tissus et cellules issus du corps humain, organes, sang et ses composants et ses produits dérivés) préalablement collectés dans le cadre du soin ou d'une recherche préalable doivent être préparés et conservés sous la forme d'une « collection d'échantillons biologiques humains ». Ce terme désigne la réunion, à des fins scientifiques, de prélèvements biologiques effectués sur un groupe de personnes identifiées et sélectionnées en fonction des caractéristiques cliniques ou biologiques d'un ou plusieurs membres du groupe, ainsi que des dérivés de ces prélèvements (article L. 1243-3 du code de la santé publique).

Cette collection est constituée en vue d'un « programme de recherche » terme désignant un ensemble d'activités de recherche organisées en vue de faciliter et d'accélérer les découvertes dans un domaine scientifique déterminé, défini par un organisme exerçant des activités de recherche ou en assurant la promotion (article L. 1243-3 du code de la santé publique).

2.3.2. Les formalités applicables

L'organisme assurant cette préparation de la collection d'échantillons médicaux doit en faire la déclaration préalable auprès du ministre chargé de la recherche (MESRI) , ainsi qu'au Directeur général de l'Agence régionale de santé lorsque l'organisme est un établissement de santé.

L'absence de respect de ce prérequis pourrait conduire à la remise en cause de la légalité du traitement ultérieur.

Cette déclaration des activités de conservation et de préparation s'effectue via l'application de gestion de la Conservation d'éléments du corps humain (CODECOH)³⁰

S'agissant des formalités à réaliser auprès de la CNIL concernant les projets de recherche :

- Si la recherche est strictement conforme à la méthodologie de référence MR-004, seule une déclaration de conformité auprès de la CNIL est nécessaire ;
- Si la recherche n'est pas conforme à ce référentiel, le responsable de traitement doit déposer une demande d'autorisation « recherche » en détaillant les points de non-conformité au référentiel correspondant. La demande devra être déposée auprès de la PDS, qui se chargera de la transmettre **au CESREES pour avis, puis à la CNIL pour autorisation.**

2.4. La protection des données à caractère personnel :

2.4.1. Champ d'application de la réglementation et principes applicables à tous les types de recherches

Le RGPD et la loi « Informatique et Libertés » s'appliquent aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches quelle que soit leur méthodologie, sous réserve qu'ils relèvent du champ d'application matériel et territorial du règlement et de la loi. La définition de la donnée à caractère personnel par le RGPD est rappelée dans le glossaire de ce document.

Il appartient alors au responsable de traitement de documenter la conformité de son projet de recherche en santé vis-à-vis du RGPD et de la loi « Informatique et libertés », notamment à l'aide des questions suivantes :

- 1) Qui est **le ou qui sont les responsable(s) de traitement dans le cadre de ce projet de recherche** ? En cas de responsabilité conjointe, un contrat conforme à l'article 26 du RGPD est-il conclu entre l'ensemble des responsables de traitement ?
- 2) Des **sous-traitants** interviennent-ils dans le projet de recherche ? Un contrat ou une convention conforme à l'article 28 du RGPD est-elle conclue ?
- 3) Quel est **l'objectif** (finalité³¹) du projet de recherche ? Cette finalité est-elle **déterminée, explicite et légitime** ? Poursuit-elle un **intérêt public**³² ?
- 4) Quelle est la **base légale**^{33, 34, 35} (**fondement juridique**) **du traitement** au sens de l'article 6 du RGPD ?
- 5) A quel titre est-il possible de **déroger au principe d'interdiction de la collecte des données de santé** dans le cadre de ce traitement au sens de l'article 9 du RGPD ³⁶ ?
- 6) Les données collectées sont-elles **adéquates, pertinentes et nécessaires** au regard de la finalité de l'étude (principe de minimisation)³⁷ ?
- 7) Les données collectées sont-elles **exactes et mises à jour** ?

³⁰ <https://www.enseignementsup-recherche.gouv.fr/fr/presentation-reglementation-conservation-elements-produits-corps-humain-fins-scientifiques-86272>

[https://entreprendre.service-public.fr/vosdroits/R17957\\$](https://entreprendre.service-public.fr/vosdroits/R17957$)

https://appliweb.dgri.education.fr/appli_web/codecoh/IdentCodec.jsp

³¹ Sur ce point, voir la fiche pratique « Définir une finalité » publiée par la CNIL.

³² Sur ce point, voir la fiche pratique « Qu'est-ce que l'intérêt public ? » publiée par le Plateforme des données de santé.

³³ Sur ce point, voir les fiches pratiques « Les bases légales » publiées par la CNIL. S'agissant des recherches cliniques, deux bases légales sont généralement privilégiées :

- l'exécution d'une mission d'intérêt public pour les structures de droit public comme les centres hospitaliers universitaires ;
- les intérêts légitimes poursuivis par le responsable de traitement pour les structures privées comme les laboratoires pharmaceutiques.

Il est à noter cependant que le Comité européen de la protection des données a retenu pour les essais cliniques de médicaments conduits selon le règlement 536/2014 et pour les traitements des données relatives à la « fiabilité » et la sécurité la base légale de « l'obligation légale » prévue au § 6.1.c) du RGPD (voir ci-dessous) ; ceci pour assurer des résultats fiables à partir de ces données, sans altération de l'intégrité par effacement ou opposition à l'utilisation.

Selon le fondement retenu, l'exercice de certains droits peut être en effet limité (droit d'opposition et droit à l'effacement) voire ne pas être applicable (droit à la portabilité).

³⁴ Direction Générale de la Santé - PP1- 9 septembre 2021, RGPD et recherches impliquant la personne humaine : clarification sur le choix de la base légale

³⁵ Pour les essais cliniques de médicaments, voir l'Avis 3/2019 du CEPD concernant les questions et réponses sur l'interaction entre le règlement relatif aux essais cliniques et le règlement général sur la protection des données (RGPD) [article 70, paragraphe 1, point b)]

³⁶ Sur ce point, l'exception prévue à l'article 9.2 (j) du RGPD permet le traitement des données de santé à des fins de recherche scientifique.

³⁷ Sur ce point, voir la fiche pratique « Minimiser les données collectées » publiée par la CNIL.

- 8) Le patient est-il **informé** au moment de la collecte conformément à l'article 69 de la loi « Informatique et libertés » et aux articles 13 ou, le cas échéant, 14 du RGPD ?³⁸ ? Peut-il exercer ses droits (accès, rectification, limitation, opposition, effacement)³⁹ ?
- 9) La **durée de conservation** des différentes catégories de données est-elle adaptée à la finalité de l'étude⁴⁰ ?
- 10) Des **mesures de sécurité**⁴¹ sont-elles mises en place, notamment avec les sous-traitants⁴², pour garantir l'intégrité, la confidentialité et la disponibilité des données conformément à l'article 32 du RGPD ?
- 11) Les **éventuels transferts de données en dehors de l'Union européenne** sont-ils encadrés conformément au chapitre V du RGPD⁴³ ?
- 12) Est-il nécessaire de réaliser **une analyse d'impact relative à la protection des données**⁴⁴ ?
- 13) Quelle est la qualification de l'étude (RIPH/RNIPH, essai clinique, investigation clinique...) ?
- 14) Est-il nécessaire de réaliser **une formalité auprès de la CNIL**⁴⁵ ?

2.4.2. Synthétique des démarches réglementaires de mise en œuvre selon la catégorie de la recherche

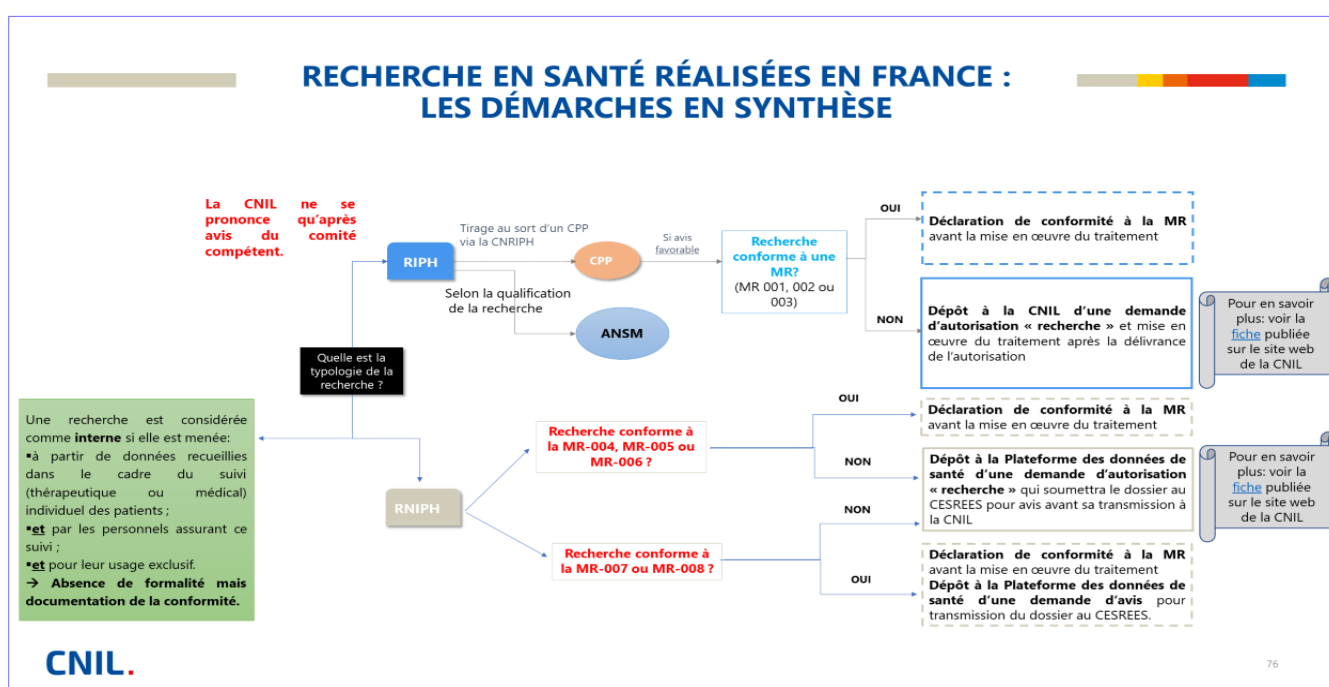


Figure 4 : les formalités applicables en fonction de la qualification de la recherche. Source : CNIL

2.4.3 L'anonymisation des données à caractère personnel

L'anonymisation est un traitement de données personnelles produisant des informations qu'il n'est pas possible, par des « moyens raisonnables », de rattacher à des personnes identifiées ou identifiables. La législation relative à la protection des données (RGPD et loi « Informatique et libertés ») reste applicable s'agissant du processus d'anonymisation mais ne s'applique plus aux données issues de ce processus dès lors que leur caractère anonyme a été démontré.

³⁸ Sur ce point, voir la [fiche pratique](#) « Traitement de données de santé : comment informer les personnes concernées ? » publiée par la CNIL ainsi que le [modèle de note d'information](#) publié par le Commission nationale des recherches impliquant la personne humaine.

³⁹ Sur ce point, voir la [fiche pratique](#) « Les droits des personnes sur leurs données » publiée par la CNIL. Le responsable de traitement doit garantir aux personnes dont les données sont traitées la possibilité d'exercer leurs droits d'accès (article 15 du RGPD), de rectification (article 16 du RGPD), d'opposition (article 21 du RGPD), d'effacement (article 17 du RGPD), de limitation (article 18 du RGPD) et de portabilité (article 20 du RGPD) dans les conditions prévues par le RGPD et la loi « Informatique et libertés ».

⁴⁰ Sur ce point, voir le [référentiel](#) sur les durées de conservation en matière de recherches dans le domaine de la santé publié par la CNIL.

⁴¹ Sur ce point, voir le [guide](#) de la sécurité des données personnelles publié par la CNIL.

⁴² Sur ce point, voir la [fiche pratique](#) « Sécurité : Gérer la sous-traitance » publiée par la CNIL.

⁴³ Sur ce point, voir la [rubrique](#) « Transférer des données hors de l'UE » sur le site web de la CNIL.

⁴⁴ Sur ce point, voir les [fiches pratiques](#) concernant l'analyse d'impact relative à la protection des données publiées par la CNIL.

⁴⁵ Sur ce point, voir la [fiche pratique](#) « Recherches dans le cadre de la santé : quelles sont les formalités ? » publiée par la CNIL.

En 2014⁴⁶, les autorités de protection des données européennes ont dégagé trois critères qui, s'ils sont conjointement réunis, permettent de conclure que les données concernées sont anonymes :

1. l'individualisation : il ne doit pas être possible d'isoler un individu dans le jeu de données ;
2. la corrélation : il ne doit pas être possible de relier les données avec un autre jeu de données distinct concernant un même individu ;
3. l'inférence : il ne doit pas être possible de déduire, de façon quasi certaine, de nouvelles informations sur un individu.

Si ces trois critères ne sont pas parfaitement remplis, le responsable de traitement qui souhaite rendre un jeu de données anonyme doit démontrer, via une évaluation approfondie des risques d'identification, que la probabilité d'identification d'une personne à l'aide de moyens raisonnablement susceptibles d'être utilisés, par lui-même ou par toute autre personne, est négligeable

Ces lignes directrices sont en cours de mise à jour au niveau européen.

S'agissant plus spécifiquement des modèles d'intelligence artificielle (IA), le CEPD a publié un avis en décembre 2024⁴⁷. Le CEPD considère que les modèles d'IA entraînés à partir de données personnelles ne peuvent pas par principe être considérés comme anonymes, et que cette appréciation devrait se faire au cas par cas.

Dans cet avis, il notamment précisé que pour considérer un modèle d'IA comme anonyme « par des moyens raisonnables », le responsable de traitement doit montrer que :

- La vraisemblance d'une extraction directe de données personnelles concernant les individus dont les données se trouvent dans la base d'entraînement est négligeable ;
- La vraisemblance d'obtenir ces données à partir de requêtes du modèle en prenant en compte les réutilisations et divulgations involontaires est négligeable.

Il a également précisé les critères d'évaluation de la vraisemblance résiduelle de réidentification (privacy by design, résistance aux attaques, documentation, etc.). Si cette liste est non exhaustive et non prescriptive, ces critères peuvent être pris en compte par les autorités pour évaluer les preuves du caractère anonyme d'un modèle par son responsable de traitement.

3. Les questions posées et les perspectives relatives aux nouvelles méthodologies d'étude

3.1. Les études combinant des catégories d'études classiques

Procédure réglementaire de mise en œuvre et comité compétent

Ces recherches sont des essais cliniques se distinguant du modèle classique de l'essai clinique contrôlé randomisé (en double aveugle ou non) par le recours à des modélisations et simulations, via l'intégration de données virtuelles, le recours à des outils statistiques avancés, et enfin l'enrichissement de sources de données avec le recours aux données de vie réelle/soins courants. Dans ces modèles, la même recherche implique donc des personnes humaines et réutilise des données existantes. Cette recherche peut de plus impliquer de nouvelles analyses sur des échantillons biologiques préalablement collectés.

Comme précisé dans l'introduction, les catégories d'études « classiques » sont les recherches organisées et pratiquées sur des personnes, les recherches avec réutilisation de données préalablement recueillies dans le cadre du soin ou d'une recherche antérieure et les recherches portant sur des échantillons préalablement prélevés et collectés. Ces trois réglementations présentent des interlocuteurs différents tant du côté opérationnel que pour les autorités compétentes, des procédures et des délais différents, un encadrement soit uniquement national soit également européen. Elles ne prévoient pas, à ce jour, de dispositions spécifiques adaptées pour encadrer les caractéristiques de ces nouvelles méthodologies d'essais.

En complément du RGPD et de la loi « Informatique et libertés » applicable à tout traitement de données à caractère personnel, il convient donc de faire application :

⁴⁶ Sur ce point, voir la [fiche pratique](#) « L'anonymisation des données personnelles » publiée par la CNIL et le chapitre 4 du livre blanc « [Données de santé artificielles : analyse et pistes de réflexion](#) » coordonnée par S. ALLASSONNIERE et J.L. FRAYSSE

⁴⁷ Sur ce point, voir l'[avis du CEPD](#) « EDPB opinion on AI models: GDPR principles support responsible AI » et la [fiche pratique](#) publiée à ce sujet « Modèles d'IA et RGPD : le CEPD publie son avis pour une IA responsable » sur le site web de la CNIL.

1. Du cadre juridique applicable au bras ou à la composante conduite sur la personne humaine « interventionnel » de l'étude (dispositions du code de la santé publique applicables aux recherches impliquant la personne humaine, règlements européens sur les essais cliniques, les dispositifs médicaux, les diagnostic in vitro, etc.) ;
2. Du cadre juridique applicable au bras ou à la composante conduite sur les données réutilisées,
3. Le cas échéant, de certaines dispositions spécifiques issues du code de la santé publique encadrant la réutilisation d'échantillons déjà collectées à des fins d'alimentation de tout ou partie du bras de contrôle externe.

Les perspectives induites par le projet de loi de simplification de la vie économique

Au regard de la législation actuellement en vigueur, les promoteurs n'ont d'autre option que de procéder au découpage « artificiel » de leur projet de recherche afin que le « groupe contrôle / témoin » ne soit soumis qu'aux dispositions du RGPD et de la loi « Informatique et libertés ». Cette solution conduit à une complexification administrative (recueil de l'avis du comité de protection des personnes ainsi que, le cas échéant, du CESREES lorsque le traitement n'est pas conforme à une méthodologie de référence de la CNIL).

Plus encore, elle peut conduire à des biais méthodologiques majeurs, l'appréciation de la pertinence scientifique de ces essais hybrides devant être évaluée dans son intégralité par un seul et unique comité de protection ou d'éthique.

Afin de remédier à ce découpage « artificiel », il a été proposé dans le projet de loi de simplification de la vie économique de compléter l'article L. 1122-1-1 par la modification suivante : « lorsque cette recherche, dans l'une de ses composantes, prévoit la réutilisation de données de santé préalablement recueillies dans le cadre de la prise en charge habituelle du participant ou d'une précédente étude, le protocole présenté à l'avis du comité de protection des personnes prévoit, pour cette composante, que les personnes sont informées conformément aux dispositions des articles 69 et 86 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et qu'elles peuvent exercer leurs droits, y compris d'opposition à la réutilisation de leurs données, conformément aux dispositions de cette même loi et du règlement général sur la protection des données (RGPD). » Cet amendement a été enregistré par le Sénat le 24 mai 2024 sous le n° COM-216.

Dans l'hypothèse où ce projet de modification législative aboutirait, il serait proposé que le comité de protection des personnes puisse se prononcer sur l'entièreté du projet. Dans cette optique, il serait toutefois particulièrement précieux que le comité de protection des personnes puisse saisir, si besoin, le CESREES conformément aux dispositions de l'article L.1123-7 du code de la santé publique et bénéficier de l'appui d'experts, lorsque cela est nécessaire, par exemple en matière de traitements de données du Système national des données de santé.

Ce sujet a fait l'objet d'une première communication au niveau européen par la France (DGS) mais doit être appuyé auprès de la Commission européenne, de ses groupes de travail, du Réseau des Chefs d'Agences (HMA) et de l'EMA, afin que les règlements encadrant les recherches sur les médicaments, les DM/DMDIV intègrent ces essais présentant une composante constituée de données de santé préalablement recueillies.

La stabilisation du cadre juridique pour l'ensemble des recherches cliniques permettra aux promoteurs de mettre en œuvre davantage d'essais hybrides. Cette typologie de cas d'usage permettra d'alimenter les travaux de la CNIL, notamment dans le cadre de la mise à jour de ses méthodologies de référence.

3.2. Les études utilisant des outils basés sur l'intelligence artificielle

Publié en juillet 2024, le règlement (UE) 2024/1689⁴⁸ (dit règlement IA (RIA ou IA Act)) établit un cadre harmonisé pour l'utilisation de l'intelligence artificielle (IA) au sein de l'Union Européenne⁴⁹. Son article 113 prévoit une entrée en application échelonnée jusqu'au 2 août 2027, date où il deviendra applicable aux systèmes d'intelligence artificielle (systèmes d'IA⁵⁰) à haut risque incorporés dans certains produits réglementés tels que les dispositifs médicaux, y compris de diagnostic *in vitro*. Il n'a pas vocation à remplacer les autres règlements (RGPD, règlement sur les DM, y compris les DM-DIV, etc.)⁵¹.

Ce règlement encadre le développement, la mise sur le marché et l'utilisation de systèmes d'IA qui peuvent poser des risques pour la santé, la sécurité ou les droits fondamentaux.

⁴⁸ Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle) (Texte présentant de l'intérêt pour l'EEE)

⁴⁹ Sur ce point, voir la [fiche pratique](#) « Entrée en vigueur du règlement européen sur l'IA : les premières questions-réponses de la CNIL » publiée par la CNIL.

⁵⁰ Sur ce point, voir les [lignes directrices](#) au sujet de la définitions des systèmes d'IA publiées par la Commission européenne.

⁵¹ Sur l'articulation entre le RGPD et le règlement sur l'IA, voir les premières [fiches pratiques](#) publiées par la CNIL.

D'un point de vue matériel, les dispositions de l'article 2 du règlement sur l'IA prévoient qu'il s'applique, par principe, aux systèmes d'IA et aux modèles d'IA à usage général. Certaines exceptions sont toutefois prévues. Il n'est ainsi pas applicable aux systèmes d'IA ou aux modèles d'IA spécifiquement développés et mis en service uniquement à des fins de recherche et développement scientifiques, ni à leurs sorties. Le considérant 25 du règlement précise à cet égard qu'« il est nécessaire de veiller à ce que le présent [...] n'affecte pas autrement les activités de recherche et de développement scientifiques relatives aux systèmes ou modèles d'IA avant leur mise sur le marché ou leur mise en service. En ce qui concerne les activités de recherche, d'essai et de développement axées sur les produits, relatives aux systèmes ou modèles d'IA, les dispositions du présent règlement ne devraient pas non plus s'appliquer avant la mise en service ou la mise sur le marché de ces systèmes et modèles. Cette exclusion est sans préjudice de **l'obligation de se conformer au [...] règlement lorsqu'un système d'IA relevant du champ d'application du [...] règlement est mis sur le marché ou mis en service à la suite de cette activité de recherche et de développement**, et sans préjudice de l'application des dispositions relatives aux bacs à sable réglementaires de l'IA et aux essais en conditions réelles. En outre, **sans préjudice de l'exclusion des systèmes d'IA spécifiquement développés et mis en service aux seules fins de la recherche et du développement scientifiques, tout autre système d'IA susceptible d'être utilisé pour mener une activité de recherche et de développement devrait rester soumis aux dispositions du [...] règlement.** »

S'agissant de son champ d'application territorial, il s'applique :

- à tous les **systèmes d'IA introduits dans le marché de l'Union européenne (UE)**, indépendamment du lieu d'établissement des opérateurs des systèmes d'IA ;
- lorsque les **données de sortie générées par le système d'IA sont utilisées dans l'UE**, même si le fournisseur ou le déployeur est situé en dehors de l'UE.

Le règlement sur l'IA prévoit une classification des systèmes d'IA en fonction des risques qu'ils comportent, notamment en cas d'atteinte aux intérêts publics et aux droits fondamentaux protégés par le droit de l'Union.

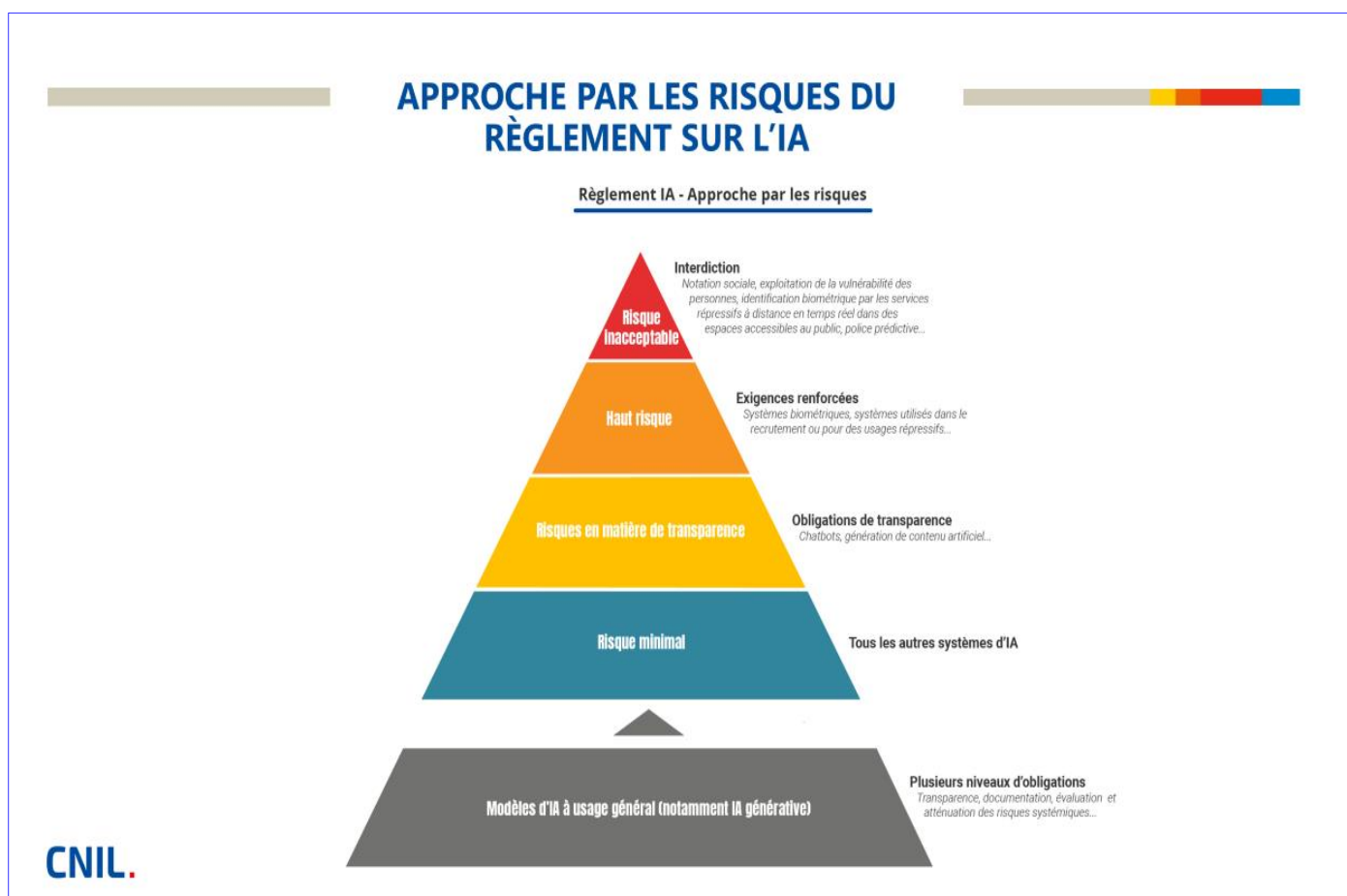


Figure 5 : Représentation de l'approche par les risques décrite dans le règlement européen encadrant l'intelligence artificielle - Source : CNIL

Sont ainsi distingués :

- **Les systèmes d'IA** mettant en œuvre des pratiques interdites et contraires aux valeurs de l'UE et aux droits fondamentaux présentant un **risque inacceptable**. Parmi les systèmes d'IA interdits listés à l'article 5 du règlement sur l'IA figurent la notation sociale, l'exploitation de la vulnérabilité des personnes, le recours à des techniques subliminales, l'utilisation par les services répressifs de l'identification biométrique à distance en temps réel dans des espaces accessibles au public, la police prédictive ciblant les individus, reconnaissance des émotions sur le lieu de travail et dans les établissements d'enseignement.
- **Les systèmes d'IA** considérés comme à **haut risque** lorsqu'ils peuvent porter atteinte à la sécurité des personnes ou à leurs droits fondamentaux, justifiant ainsi que leur développement soit soumis à des exigences renforcées. Ces systèmes sont listés dans l'annexe I du règlement pour les systèmes intégrés dans des produits qui font déjà l'objet d'une surveillance de marché (dispositifs médicaux, jouets, véhicules, etc.) et dans l'annexe III pour les systèmes utilisés dans huit domaines spécifiques (article 6 du règlement sur l'IA).
- **Les autres systèmes d'IA parmi lesquels :**
 - o **les systèmes d'IA pouvant être à l'origine de trucage ou de manipulation** dont les fournisseurs et les déployeurs sont **soumis à des obligations de transparence spécifiques**. Il en va ainsi notamment des systèmes destinés à interagir directement avec les personnes physiques et ceux qui génèrent des contenus de synthèse (article 50 du règlement sur l'IA) ;
 - o **les systèmes d'IA** présentant un **risque minimal ou nul** pour lesquels le règlement ne prévoit pas d'obligation spécifique.

Le règlement sur l'IA prévoit également des exigences distinctes en fonction de la qualité de l'acteur (fournisseur, importateur, distributeur, déployeur et fabricant) dans le cycle de vie l'IA pour un système d'IA donné.

S'agissant des systèmes d'IA à haut risque, les articles 8 et suivants du règlement sur l'IA prévoient les exigences qui leur sont applicables, dont :

- La mise en place d'un **système de gestion des risques documenté et mis à jour**. Ce système, qui comporte un certain nombre d'étapes est défini comme « un processus itératif continu qui est planifié et se déroule sur l'ensemble du cycle de vie d'un système d'IA à haut risque et qui doit périodiquement faire l'objet d'un examen et d'une mise à jour méthodiques ». Plusieurs éléments doivent être pris en compte dans ce système de gestion des risques conformément à l'article 9 du règlement sur l'IA ;
- La mise en place d'une **gouvernance et d'une gestion des données appropriée** pour s'assurer que les jeux de données d'entraînement, de validation et de test sont de qualité suffisante pour éviter les biais (article 10 du règlement sur l'IA) ;
- L'établissement d'une **documentation technique** avant la mise sur le marché ou la mise en service pour démontrer que le système d'IA à haut risque satisfait aux exigences énoncées dans la section 2 du chapitre III du règlement afin de « fournir aux autorités nationales compétentes et aux organismes notifiés les informations nécessaires sous une forme claire et intelligible pour évaluer la conformité du système d'IA avec ces exigences » (article 11 du règlement sur l'IA). La durée de conservation de cette documentation technique est fixée par le règlement (article 18) ;
- La mise en place de systèmes d'enregistrement des événements (**journaux d'événements**) (articles 12 et 19 du règlement sur l'IA) ;
- La **transparence** quant au système d'IA et à son fonctionnement grâce à la fourniture d'une notice d'utilisation aux déployeurs du système afin de leur permettre « d'interpréter les sorties d'un système et de l'utiliser de manière appropriée » (article 13 du règlement sur l'IA)⁵² ;
- La mise en place d'un **contrôle humain**, notamment par le développement « d'interfaces homme-machine appropriées pour permettre un contrôle effectif de leur utilisation par des personnes physiques » afin de « prévenir ou réduire au minimum les risques pour la santé, la sécurité ou les droits fondamentaux qui peuvent apparaître

⁵² En droit français, les dispositions de l'article L.4001-3 du code de la santé publique encadrent l'usage, pour un acte de prévention, de diagnostic ou de soin, de dispositifs médicaux comportant un traitement de données algorithmique dont l'apprentissage a été réalisé à partir de données massives. Dans cette hypothèse, le professionnel de santé doit informer la personne concernée de l'usage d'un traitement algorithmique et, le cas échéant, de l'interprétation qui en résulte. Par ailleurs, le professionnel de santé doit également être informé du traitement de données (et notamment des données utilisées et des résultats). Enfin, les concepteurs du traitement algorithmique doivent s'assurer de l'explicabilité de son fonctionnement pour les utilisateurs. La nature des dispositifs médicaux concernés et leurs modalités d'utilisation doivent être précisés par un arrêté du ministre chargé de la santé établi après avis de la Haute autorité de santé et de la CNIL. Il n'a, pour l'heure, pas été publié.

lorsqu'un système d'IA à haut risque est utilisé conformément à sa destination ou dans des conditions de mauvaise utilisation raisonnablement prévisible » (article 14 du règlement sur l'IA) ;

- La mise en place de mesures afin d'atteindre **un niveau approprié d'exactitude, de robustesse et de cybersécurité** tout au long du cycle de vie du système d'IA (article 15 du règlement sur l'IA) ;
- La mise en place d'un **système de gestion de la qualité pour assurer la conformité à toutes les étapes du cycle de vie du produit**, y compris la surveillance après sa commercialisation (article 17 du règlement sur l'IA) ;
- La réalisation d'une **analyse d'impact sur les droits fondamentaux** que le système d'IA peut produire (article 27 du règlement sur l'IA) ;
- L'obtention d'un marquage CE ;
- La rédaction d'une **déclaration de conformité** (article 47 du règlement sur l'IA) ;
- L'enregistrement dans la **base de données de l'UE** avant la mise sur le marché (article 71 du règlement sur l'IA) ;
- La notification des **incidents graves** aux autorités de surveillance des marchés des Etats membres dans lesquels l'incident s'est produit (article 73 du règlement sur l'IA).

A l'instar du règlement EEDS, le règlement sur l'IA nécessitera une adaptation du droit français. Par ailleurs, son entrée en application s'appuiera sur des « **normes harmonisées** » au niveau européen qui doivent définir précisément les exigences applicables aux systèmes d'IA concernés. Dans ce cadre, la Commission européenne a commandé au CEN/CENELEC (Comité européen de normalisation en électronique et en électrotechnique) dix normes actuellement en cours de rédaction.

Cette annexe souligne la difficulté croissante d'appliquer un cadre juridique sectoriel spécifique pour chaque catégorie de recherche alors même que, de plus en plus, celles-ci croisent intervention sur la personne, données secondaires, modélisation et intelligence artificielle. Ces méthodologies hybrides, aujourd'hui au cœur de nombreuses innovations en recherche clinique, bousculent les frontières traditionnelles entre les types d'études et révèlent les limites d'un système normatif fragmenté.

L'empilement de réglementations nationales et européennes, avec des autorités, des procédures et des délais propres à chaque composante, crée des parcours complexes pour les promoteurs. Cette hétérogénéité n'est pas seulement un enjeu administratif : elle pose un risque de perte de cohérence méthodologique, entrave l'évaluation scientifique intégrée des projets et peut ralentir l'accès à des approches pourtant prometteuses pour la santé publique.

Dans ce contexte, il devient nécessaire de renforcer la lisibilité du cadre applicable aux études combinées, en favorisant une articulation fluide des dispositifs existants et une reconnaissance explicite des modèles hybrides. L'encadrement des systèmes d'IA, la réutilisation des données de santé et l'intégration de composantes non interventionnelles ne doivent plus être pensés comme des ajouts périphériques, mais comme des éléments à part entière de la conception d'un essai.

Assurer la qualité, la pertinence et la robustesse de ces recherches implique de conjuguer sécurité juridique, protection des personnes et pertinence scientifique. Ce travail d'harmonisation réglementaire et de clarification méthodologique est indispensable pour permettre à ces nouveaux formats d'étude de s'inscrire durablement dans l'arsenal de la recherche clinique, et d'en renforcer l'impact au service de l'innovation en santé.

Annexe 3 - Cross validation des solutions proposées avec les recommandations émises par EMA sur les essais non contrôlés « monobras »

Type de biais	Description	Possibles mesures de réduction du biais	Méthodologie(s) et outils statistiques susceptible(s) de limiter ce biais
Biais de détermination	Les données externes varieront probablement selon certains aspects de la conduite des essais et de la collecte des données, p. ex., la fréquence et les normes d'évaluation des résultats diffèrent de celles des données recueillies dans le cadre de l'essai monobras	Lorsqu'on utilise des données externes à l'étude, les pratiques de vérification et de collecte de données doivent être comparables à celles de l'étude expérimentale.	L'émulation d'essais ciblés spécifie clairement les critères d'éligibilité, les interventions, et les critères de jugement à l'avance, réduisant ainsi les risques de biais de détermination et d'évaluation grâce à l'utilisation de définitions opérationnelles standardisées et validées pour mesurer les expositions et les résultats Anticipation d'un recueil de données de façon prospective.
Biais d'évaluation	La connaissance du traitement peut influencer l'évaluation des résultats.	Les critères d'évaluation des essais doivent être suffisamment objectifs et, si possible, les évaluations doivent être effectuées de manière indépendante et de préférence sans connaître le moment du traitement.	
Biais d'attrition	L'attrition des patients et les données manquantes en général constituent une source supplémentaire de confusion difficile à résoudre.	Éviter les données manquantes grâce à la conception et à la conduite de l'étude. Pré-spécifier les méthodes de traitement des données manquantes qui ne surestiment pas la réponse au traitement et effectuer des analyses de sensibilité appropriées. Lors de l'utilisation de données externes, les données doivent être de haute qualité avec un suivi incluant tous les patients, afin d'éviter les biais dus aux données manquantes.	En définissant des critères de suivi précis et en appliquant des méthodes d'analyse qui prennent en compte l'attrition (par exemple, l'analyse selon le principe de l'intention de traiter), l'approche d'émulation cherche à minimiser les effets de l'attrition. Anticipation d'un recueil de façon prospective.
Biais dû à l'absence de planification préalable	Toute modification de la conception, de la conduite et des rapports prévus après le début de l'essai (p. ex., plan d'analyse statistique, adaptations du traitement, suivi, modifications du protocole sur les critères	La planification préalable est essentielle pour tous les essais de confirmation, mais le standard doit être encore plus élevé pour les essais monobras (p. ex., le plan d'analyse statistique doit être finalisé avant le début de l'essai, il faut absolument minimiser les changements du	En exigeant un protocole d'étude pré-spécifié qui simule un essai randomisé, y compris les plans d'analyse statistique, l'approche d'émulation aborde directement ce biais. Anticipation d'un recueil de façon prospective.

	d'inclusion ou d'exclusion, traitement concomitant autorisé) comporte le risque d'introduire un biais.	protocole et du plan d'analyse statistique après le début de l'essai, Si des analyses intermédiaires sont planifiées, il peut exister des problèmes si elles sont flexibles ou non effectuées lors de l'atteinte du niveau d'information prévu).	
Biais dû à la régression à la moyenne	Les patients sélectionnés en fonction de leurs valeurs de résultat pendant la période d'observation peuvent présenter de meilleurs résultats en raison de la régression de la moyenne	Définir la population cible indépendamment de la gravité de la maladie pendant la période d'observation préalable au traitement. Éviter de sélectionner les patients en fonction de mesures de résultats sujettes à des erreurs de mesure ou à des fluctuations.	Les plans d'analyse dans l'émulation d'essais ciblés peuvent inclure des ajustements pour les mesures de base et les caractéristiques cliniques pour gérer la régression vers la moyenne et la variabilité dans l'histoire de la maladie Anticipation d'un recueil de façon prospective.
Biais dû à la variabilité de l'histoire de la maladie	Les patients peuvent avoir une variabilité importante dans leur historique de la maladie avant que le médicament expérimental ne soit administré. Ceci est particulièrement (mais pas seulement) préoccupant pour les paramètres de temps de survenu d'événement où l'historique de la maladie est généralement fortement lié et pronostique du résultat individuel.	L'analyse des paramètres temps jusqu'à événement est généralement plus difficile à évaluer sans biais. Les paramètres et les méthodes d'analyse qui ne reposent pas directement sur une échelle de temps devraient être choisis.	
Biais de calendrier	Le traitement standard et les tendances dans la gestion globale de la maladie peuvent changer l'évolution de la maladie et les résultats individuels au fil du temps. Dans les essais monobras, l'impact de ces tendances ne peut être dissocié de l'effet du traitement.	Utilisation de données externes contemporaines.	Lors de la réalisation de l'essai émulé, vérifier la contemporanéité des données. Le recueil prospectif ne permet pas de répondre à ce biais mais dans la mesure où la randomisation est possible, ce critère devient peu pertinent.
Biais d'immortalité	Le début de l'étude ou le début du traitement pour les études précédentes ou les données externes est difficile à déterminer en tant que point de départ d'une échelle de temps spécifique au patient.	Le temps de début de l'exposition au risque (T0) doit être clairement définie, si des critères d'évaluation de type temps jusqu'à événement doivent être utilisés et comparés. De plus, des analyses de sensibilité sont nécessaires.	L'approche d'émulation utilise des méthodes de conception et d'analyse, comme la synchronisation des débuts de suivi entre les 2 bras, les modèles de risques proportionnels de Cox pour gérer correctement le biais d'immortalité. Il faut donc identifier ou recueillir l'élément synchronisateur
Biais lié aux événements intercurrents survenant après le début de l'étude	Absence de définition claire des estimands et des événements intercurrents d'intérêt lors de la planification de l'essai.	En suivant les recommandations de l'ICH E9 (R1), anticiper les événements intercurrents d'intérêt dès la conception de l'essai et s'assurer de la bonne définition des estimands	Anticipation d'un recueil de façon prospective.
Biais de sélection rétrospective	La sélection rétrospective de données externes à utiliser comme référence et la	Spécifier préalablement au début de l'essai l'utilisation de données externes, ainsi que les détails de l'analyse	L'approche d'émulation minimise ces biais par la spécification a priori des critères d'éligibilité et des méthodes d'assignation des expositions

	spécification des principales caractéristiques d'analyse a posteriori peuvent introduire un risque de biais.	statistique. Le plan d'analyse statistique doit être finalisé avant le début de l'essai.	basées sur ces critères, ainsi que l'utilisation de méthodes d'appariement ou de scores de propension pour assurer l'équilibre entre les groupes Anticipation d'un recueil de façon prospective.
Biais de sélection lié au groupe témoin hypothétique	Les patients enrôlés dans un essai monobras peuvent différer systématiquement du groupe témoin hypothétique sur des facteurs pouvant avoir une incidence sur leur pronostic.	Préciser a priori clairement les critères d'inclusion et d'exclusion de manière à ce que la population enrôlée dans l'essai corresponde bien aux données externes sur lesquelles reposent les hypothèses.	L'approche d'émulation minimise ces biais par la spécification a priori des critères d'éligibilité et des méthodes d'assignation des expositions basées sur ces critères, ainsi que l'utilisation de méthodes d'appariement ou de scores de propension pour assurer l'équilibre entre les groupes Anticipation d'un recueil de façon prospective.
Biais de sélection lié à la population cible	Les patients enrôlés dans un essai monobras peuvent différer systématiquement de la population cible sur des facteurs ayant un impact sur leur pronostic.	Limiter le nombre et l'étendue des critères d'inclusion et d'exclusion. Pré-spécifier clairement le pronostic attendu en termes de critère principal de la population cible, incluant les données externes sur lesquelles il est basé	L'approche d'émulation minimise ces biais par la spécification a priori des critères d'éligibilité et des méthodes d'assignation des expositions basées sur ces critères, ainsi que l'utilisation de méthodes d'appariement ou de scores de propension pour assurer l'équilibre entre les groupes Anticipation d'un recueil de façon prospective.
Biais de sélection lié aux sous-groupes définis par biomarqueurs	Les patients sélectionnés sur la base d'un biomarqueur prédéfini pour le traitement ciblé peuvent différer dans le pronostic par rapport à la population totale.	Veiller à ce que le pronostic du sous-groupe de patients identifié par le biomarqueur soit connu avec une précision suffisante avant l'essai	Anticipation d'un recueil de façon prospective. Utilisation de biobanques pour compléter a posteriori quand le test sera disponible.
Biais de migration de stade	L'amélioration des méthodes d'évaluation conduit à une amélioration du pronostic des stades précoces et tardifs.	Veiller à ce que les mêmes méthodes d'évaluation soient utilisées dans l'essai monobras et dans les sources de données externes ; Démontrer une ampleur d'effet qui supérieur à l'impact maximal possible de l'effet de migration de stade.	Anticipation d'un recueil de façon prospective.
Biais d'étude	Les patients dans un essai monobras peuvent avoir des résultats systématiquement différents (indépendamment du traitement expérimental) que dans la pratique clinique ciblée, p. ex., en raison de soins différents dans le cadre de l'essai.	Vérifier et montrer que les soins annexes reflètent le standard de soins actuel	En simulant la rigueur d'un essai randomisé, y compris la pré-spécification des plans d'analyse et la considération des standards éthiques, l'approche d'émulation vise à minimiser le biais global de l'étude. Anticipation d'un recueil de façon prospective.

Annexe 4 - Analyse des recommandations internationales sur la simulation in silico et l'intelligence artificielle

Au cours des dernières années, les agences réglementaires ont reconnu l'importance croissante de l'intelligence artificielle (IA) dans le développement, l'évaluation et la surveillance des produits de santé. Face aux défis posés par l'intégration de ces technologies dans le cycle de vie du médicament, plusieurs organismes ont publié des recommandations visant à encadrer leur utilisation. Aux États-Unis, la FDA a mis en place le SaMD Action Plan et a publié début 2025 les « *considerations for the use of AI to support regulatory decision making for drug and biological product* ». En Europe, l'EMA a publié en septembre 2024 un « *Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle* » détaillant les bonnes pratiques pour l'application de l'IA dans la recherche et les essais cliniques, avec un cadre spécifique pour la qualification des méthodologies algorithmiques.

A l'échelle internationale, l'ICH M15 établit des principes harmonisés pour la validation des modèles prédictifs, assurant une approche cohérente entre les différents régulateurs.

Ces différentes sources ont fait l'objet d'une analyse afin d'identifier les principaux points soulevés, mettant en lumière les exigences en matière de transparence, de validation, et de surveillance post-commercialisation des systèmes IA dans le domaine pharmaceutique et sont résumées dans le tableau ci-après, complété par les références des différents sources évoquées.

Description	Informations détaillées requises	Recommandations et Réglementation Applicables
1. Objectif et portée		
Définit comment le modèle sera utilisé dans le développement du produit médical et son évaluation réglementaire. Doit inclure le but du modèle et la portée de son utilisation.	<ul style="list-style-type: none"> - Application clinique (question of interest) : détailler l'utilisation prévue dans le développement (endpoint, prédiction de réponse, stratification) et alignée sur le contexte réglementaire. Le modèle doit répondre à une question spécifique d'intérêt. - Contexte d'utilisation (context of use) : Intégrer les données IA/ML dans des essais cliniques ou en routine clinique. - Bénéfices attendus : Justifier la valeur ajoutée clinique et opérationnelle par rapport aux méthodes conventionnelles. - Evaluer l'impact réglementaire (risque) 	<p>EMA - Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle (2024)⁵³</p> <p>FDA – Draft - Considerations for the use of AI to support regulatory decision making for drug and biological product (2025) ⁵⁴</p>
2. Qualité des données		
Population cible	<ul style="list-style-type: none"> - Caractéristiques cliniques et démographiques : Décrire les patients inclus dans les essais (âge, sexe, comorbidités). - Variabilité inter-individuelle : Documenter la manière dont la variabilité clinique et biologique est prise en compte. - Critères d'exclusion : Identifier les groupes pour lesquels le modèle n'est pas validé. - Représentativité : Justifier que la population étudiée reflète celle rencontrée en pratique clinique ou dans les essais cliniques. Vérifier que les données incluent des sous-groupes minoritaires. - Impact clinique : Expliquer comment l'IA améliore la prise en charge de ces sous-populations. 	<p>EMA - Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle (2024)</p> <p>FDA – Draft - Considerations for the use of AI to support regulatory decision making for drug and biological product (2025)</p>
Données d'apprentissage	<ul style="list-style-type: none"> - Origine et diversité des données : Intégrer des données cliniques, biologiques et in silico provenant de registres ou d'essais. - Prétraitement : Détailler les étapes (nettoyage, anonymisation, normalisation, transformation et augmentation des données). - Contrôle des biais : Documenter les procédures visant à minimiser les biais dans les données, notamment le biais d'échantillonnage (détecter et corriger les déséquilibres, par ex., données rares). 	<p>EMA - Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle (2024)</p> <p>FDA – Draft - Considerations for the use of AI to support regulatory decision making for drug and biological product (2025)</p> <p>RGPD - Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016⁵⁵</p> <p>ICH - Guideline for good clinical practice (GCP) E6(R3) E6⁵⁶</p>

⁵³ https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf

⁵⁴ <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological>

⁵⁵ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

⁵⁶ <https://www.ema.europa.eu/en/ich-e6-good-clinical-practice-scientific-guideline>

	- Sécurisation des données : Préciser les pratiques garantissant l'intégrité et la confidentialité.	
3- Apprentissage et performance		
Méthodologie d'apprentissage	<ul style="list-style-type: none"> - Type d'apprentissage : Non supervisé, semi-supervisé, supervisé, par renforcement. - Techniques avancées : Mentionner l'usage de techniques spécifiques (ex. : IA générative, apprentissage fédéré). - Paramètres critiques : Identifier les hyperparamètres influençant les performances. - Choix des algorithmes : Expliquer les critères pour sélectionner les modèles utilisés. - Validation croisée : Utiliser des données indépendantes pour éviter le surapprentissage et décrire la méthodologie utilisée pour garantir la généralisabilité du modèle. 	<p>EMA - Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle (2024)</p> <p>ICH- M15 Draft Guideline (2024) – General principles for model-informed drug development (MIDD)⁵⁷</p> <p>ISO - Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)⁵⁸</p>
Performance et validation Quantifie les erreurs numériques dues aux approximations dans le modèle.	<ul style="list-style-type: none"> - Métriques clés : Fournir des indicateurs robustes et adaptés au contexte (ex. : AUC, RMSE, R², sensibilité, spécificité) - Comparaison à l'étalon-or : Comparer les performances aux standards existants ou au jugement d'experts. 	<p>EMA - Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle (2024)</p> <p>FDA - considerations for the use of AI to support regulatory decision making for drug and biological product</p> <p>FDA - SaMD Action Plan (2021)</p> <p>ICH E9 : Lignes directrices pour la définition des endpoints dans les essais cliniques⁵⁹</p>
Robustesse, Surveillance et dérives	<ul style="list-style-type: none"> - Concept drift : Mettre en place des mécanismes pour détecter les dérives conceptuelles ou les dégradations de performance. - Analyse des valeurs extrêmes (outliers) : Expliquer les méthodologies pour traiter les valeurs atypiques. - Robustesse : Tester les performances dans des scénarios variés (ex. : données bruitées, capteurs différents). - Mises à jour incrémentales : Détailler les processus pour maintenir la performance du modèle après déploiement. - Validation continue : Tester périodiquement le modèle avec des données récentes. Détailler les processus pour garantir la stabilité et la fiabilité après les modifications. 	<p>EMA - Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle (2024)</p> <p>FDA – Draft - Considerations for the use of AI to support regulatory decision making for drug and biological product (2025)</p>

⁵⁷ https://database.ich.org/sites/default/files/ICH_M15_EWG_Step2_DraftGuideline_2024_1031.pdf

⁵⁸ <https://www.iso.org/standard/74438.html>

⁵⁹ <https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline>

	- Documentation des versions : Assurer une traçabilité complète des versions de modèles utilisés dans les essais.	
4- Preuve clinique et impact		
Validation clinique et génération de preuve	<ul style="list-style-type: none"> - Validation méthodologique : décrire les études cliniques démontrant que la méthodologie/ modèle est fiable et améliore la précision ou la sensibilité des études cliniques - Reproduction des essais cliniques passés : Valider la méthodologie IA via des études dont les résultats sont déjà connus. - Inclusion du modèle /IA dans les essais prospectifs : Tester les prédictions dans des conditions réelles. - Mesures d'impact clinique : Montrer la corrélation entre les prédictions IA et les critères de jugement cliniques pertinents. - Exemples d'application : fournir des preuves issues d'essais ou de données réelles montrant la valeur ajoutée de la méthodologie 	<p>ICH M15 Draft Guideline (2024) – General principles for model-informed drug development (MIDD)</p> <p>FDA – Draft - Considerations for the use of AI to support regulatory decision making for drug and biological product (2025)</p>
Impact clinique	<ul style="list-style-type: none"> - Endpoint clinique : Décrire l'apport du modèle ou des prédictions IA dans les mesures ou évaluations de critères de jugement critiques. - Bénéfices prouvés : Quantifier les gains (temps, précision, coût) obtenus grâce à l'IA. - Concordance avec les standards : Vérifier que les résultats sont alignés avec les lignes directrices (ex. : standards EMA/FDA). - Plan de gestion des risques : Identifier les risques potentiels et fournir des stratégies pour minimiser les risques associés à l'utilisation clinique du modèle (biais, erreurs de prédiction). 	<p>ICH M15 Draft Guideline (2024) – General principles for model-informed drug development (MIDD)</p> <p>FDA – Draft - Considerations for the use of AI to support regulatory decision making for drug and biological product (2025)</p>
5- Éthique et réglementation		
Explicabilité et transparence	<ul style="list-style-type: none"> - Rapports compréhensibles : Documenter les sorties explicatives compréhensibles pour les cliniciens et les régulateurs. - Auditabilité : Préparer une documentation complète pour un audit externe. - Imputabilité des prédictions : Garantir que les décisions IA peuvent être justifiées de manière transparente. 	<p>EMA - Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle (2024)</p> <p>FDA – Draft - Considerations for the use of AI to support regulatory decision making for drug and biological product (2025)</p> <p>ISO - Guidelines on the testing of AI-based systems (2020) ⁶⁰</p>
Conformité réglementaire	- Qualification EMA/FDA : Décrire les interactions avec les agences réglementaires et les étapes de qualification ou d'approbation.	EMA - Qualification opinion for Artificial Intelligence-Based Measurement of Non-alcoholic Steatohepatitis Histology in Liver

⁶⁰ <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:tr:29119:-11:ed-1:v1:en>

<p>- Gestion éthique : Assurer une gouvernance appropriée des données (respect du consentement, anonymisation).</p> <p>- Imputabilité des parties : Identifier les rôles et responsabilités des équipes impliquées (développeurs, cliniciens, experts en IA).</p>	<p>Biopsies to Determine Disease Activity in NASH/MASH Clinical Trials, Feb 2025)⁶¹</p> <p>Food and Drug Administration (FDA): SaMD Action Plan (2021) ⁶²</p> <p>RGPD - Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016</p> <p>ISO - Medical devices — Quality management systems — Requirements for regulatory purposes (2016) ⁶³</p>
---	--

⁶¹ https://www.ema.europa.eu/en/documents/other/qualification-opinion-artificial-intelligence-based-measurement-non-alcoholic-steatohepatitis-histology-liver-biopsies-determine-disease-activity-nash-mash-clinical-trials_en.pdf

⁶² <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>

⁶³ <https://www.iso.org/fr/standard/59752.html>

