Modèle de protocole scientifique MR-004

A remplir

| ***Aide au remplissage***  *(*Cette partie est à effacer avant envoi du projet*)* |
| --- |

*Afin de vous aider à documenter les caractéristiques de votre projet de recherche mis en œuvre dans le cadre de la procédure simplifiée MR-004, nous vous mettons à disposition ce modèle de protocole scientifique. La réalisation d’un protocole scientifique est une exigence de la MR-004 : la CNIL pourra vous demander de le lui présenter en cas de contrôle. Un tel document vous sera par ailleurs demandé dans le cadre de l’enregistrement de votre projet sur le répertoire public du Health Data Hub conformément aux obligations de transparence associées à cette procédure simplifiée. Vous êtes libre de ne pas utiliser ce modèle. Nous vous conseillons d’appliquer la taille de la police standard selon les usages en vigueur (Calibri 10 minimum par exemple).*Dans la suite du protocole, nous vous invitons à effacer systématiquement les parties “Aide au remplissage” et “Nos conseils” pour ne laisser apparaître que vos réponses.

En cas de doute sur les démarches réglementaires à accomplir pour accéder aux données de santé et éventuellement l’éligibilité de votre étude à la méthodologie de référence 004, vous pouvez contacter le HDH ([hdh@health-data-hub.fr](mailto:hdh@health-data-hub.fr)) ou consulter le site du [HDH](https://www.health-data-hub.fr/demarches-reglementaires) ou de la [CNIL](https://www.cnil.fr/fr/traitements-declaration-conformite?field_norme_numerotation_type_value%5B0%5D=6).

| ***Nos conseils***  *(*RAPPEL : cette partie est à effacer avant envoi du projet*)* |
| --- |

Si vous voulez accéder à des données dont vous n’êtes pas le dépositaire, vous aurez besoin d’aide pour remplir les rubriques “3.1 - Données requises”, nous vous conseillons vivement de prendre contact avec les équipes en charge de ces données. Si vous n’allez pas héberger vous-même les données nécessaires pour votre projet, vous aurez besoin de l’aide du responsable du système d’information concerné pour remplir l’analyse d’impact relative à la protection des données (AIPD).

A toutes les étapes de la démarche, vous pouvez :

* Consulter le [répertoire des projets](https://www.health-data-hub.fr/projets) pour avoir une idée de ce qui est réalisé à partir des données ;
* Consulter notre [guide pédagogique](https://health-data-hub.fr/sites/default/files/2025-10/Guide%20p%C3%A9dagogique%20Proc%C3%A9dure%20simplifi%C3%A9e%20MR-004.pptx.pdf) sur la MR-004 et notre [checklist](https://health-data-hub.fr/sites/default/files/2025-10/202510-%20HDH_Starterkit%20MR004%20_%20Checklist%20d%27%C3%A9valuation%20de%20la%20conformit%C3%A9%20%281%29.docx) pour connaître les principales exigences de cette MR et les modalités associées pour la mettre en oeuvre ;
* Faire appel à la [communauté](https://entraide.health-data-hub.fr/) ;
* Participer aux [formations](https://health-data-hub.fr/formations) organisées par le HDH pour avoir un premier niveau d'information sur le [SNDS](https://documentation-snds.health-data-hub.fr/) et aux [meet-up](https://www.meetup.com/fr-FR/Health-Data-Hub/) pour échanger avec tout type de porteurs de projet
* Vous référer à la [documentation collaborative](https://documentation-snds.health-data-hub.fr/) ou utiliser un des programmes partagés par Ia communauté.

**[TITRE COMPLET DE LA RECHERCHE]**

**Sommaire du protocole scientifique**

[**PRÉSENTATION DE L’ÉQUIPE PROJET**](#_heading=h.30j0zll) **5**

[**OBJECTIFS ET FINALITÉS**](#_heading=h.naq8wt60i932) **7**

[Contexte, objectif(s) et justification de l’étude](#_heading=h.6fp7ife5ig56) 7

[Justification du respect de l’éthique](#_heading=h.pflbil6sfrq5) 7

[Justification de l’intérêt public](#_heading=h.9qyrc96005c4) 7

[Publication des résultats et valorisation](#_heading=h.gpdr4t8skkkt) 9

[**MÉTHODOLOGIE**](#_heading=h.adc6t9m4vg0z) **10**

[Design de l’étude](#_heading=h.42xcfldzewtk) 10

[Description et justification de la population d’étude](#_heading=h.nwwml3kzvn1s) 10

[Taille de la population](#_heading=h.oucs5ji4u46a) 12

[Sources de données](#_heading=h.ubehqw53weui) 12

[Variables](#_heading=h.al8a34qnqraf) 13

[Préparation de données](#_heading=h.ebkzk0rywe7m) 13

[Méthodes, traitements et analyses des données](#_heading=h.3i6jdop8v0gj) 14

[Limites de l’étude](#_heading=h.vxmhh5ujv63) 14

[Calendrier prévisionnel et faisabilité du projet](#_heading=h.310tmdqe3ndv) 15

[**ANNEXES**](#_heading=h.c83bjdw1yuhg) **26**

# PRÉSENTATION DE L’ÉQUIPE PROJET

| ***Aide au remplissage***  *(*RAPPEL : cette aide au remplissage est à effacer avant envoi du projet*)* |
| --- |

Désigner les structures Responsable(s) de traitement et du/des responsable(s) de mise en œuvre du projet :

* Le / les responsable(s) de traitement sont : si besoin, vous référez à la définition ci-après.
* Le / les responsable(s) de la mise en œuvre sont : si besoin, vous référez à la définition ci-après.

Présentez-les.

Indiquer le nom de l’organisme du financeur.

Désigner les représentants de ces structures et lister les membres de l’équipe projet, leur fonction, et leur rôle au sein du projet.

Désigner le Délégué à la Protection des Données. Nous vous conseillons de préciser son nom, son adresse postale, téléphone, mail, site web.

Si dans votre étude un comité scientifique est mis en place, préciser son rôle, mentionner ses membres et leurs affiliations.

| ***Définition***  *(*RAPPEL : cette partie est à effacer avant envoi du projet*)* |
| --- |

**Le responsable d’un traitement** de données à caractère personnel est la personne, l’autorité publique, le service ou l’organisme qui détermine les finalités, les objectifs et les moyens de l’étude. Il endosse la responsabilité juridique du traitement et le récépissé de déclaration de conformité à la MR-004 est établi à son attention. Voici une liste d’indices permettant d’identifier le responsable de traitement :

* Il est à l'initiative du traitement ;
* Il décide de l'objectif et des résultats du traitement ;
* Il décide quelles données vont être collectées ;
* Il décide quelle population va être concernée par le traitement ;
* Il a une totale autonomie dans la détermination des moyens et finalités du traitement ;
* Il choisit les sous-traitants pour traiter les données pour son propre compte ;
* Le responsable de traitement doit être représenté par son représentant légal.

**Le responsable de mise en œuvre** est en charge de réaliser l’étude ou une partie de l’étude ; s’il n’est pas identique au responsable de traitement, il est son sous-traitant.

**Le financeur** est l’organisme qui fournit tout ou partie des capitaux nécessaires à la réalisation de l’étude, recherche ou évaluation. S’il est identique au responsable de traitement, ce champ n’est pas à remplir.

**Le délégué à la protection des données** assure la qualité de l’information des personnes concernées par l’étude et s’assure de l'absence d’opposition de ces personnes, conformément au RGPD et la loi informatique et liberté (LIL).

**Le comité scientifique de l’étude** est un comité d’experts indépendants mis en place par l’équipe projet pour assurer le bon déroulement de l’étude d’un point de vue scientifique. Ce comité est généralement composé au minimum de 3 membres de différentes spécialités (par ex. un(e) neurologue, un(e) épidémiologiste et un(e) économiste de la santé). Le comité scientifique est facultatif et ne s’applique pas à toutes les études. Il est cependant préconisé pour les études nécessitant une expertise particulière ou ayant recours à des méthodologies innovantes.

Un **système du SNDS** est un système rassemblant des données mentionnées au I du L. 1461-1 du [code de la santé publique](https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006072665&dateTexte=&categorieLien=cid) et assurant leur mise à disposition pour les finalités prévues au III du même article. Il existe trois types de systèmes du SNDS :

* Le **SNDS central** qui est un système du SNDS réunissant, organisant et mettant à disposition l'ensemble des données qui constituent la base principale et les bases du catalogue. La responsabilité de traitement est conjointe entre la CNAM et la Plateforme des données de santé (en application de l'article R. 1461-3 du CSP).
* Les **Système fils** qui sont des systèmes du SNDS rassemblant des données mentionnées au I du L. 1461-1 du [code de la santé publique](https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006072665&dateTexte=&categorieLien=cid) provenant d'un système du SNDS (SNDS central, système fils du SNDS ou autre système du SNDS) et assurant leur mise à disposition pour les finalités prévues au III du même article.
* Les **autres systèmes du SNDS** qui sont des systèmes du SNDS, rassemblant des données mentionnées au I du L. 1461-1 du [code de la santé publique](https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006072665&dateTexte=&categorieLien=cid), ne répondant ni à la définition de SNDS central, ni à celle de système fils.

Ainsi, lorsqu’il est fait mention dans le protocole des “données du SNDS”, cela vise les données mises à dispositions par l’ensemble des systèmes du SNDS. A noter que le traitement des données du SNDS dans le cadre de la MR-004 est possible, il n’existe aucune limitation sauf lorsque les données proviennent directement de la base principale, conformément à ce qui est précisé dans le présent document.

| ***Nos conseils***  *(*RAPPEL : cette partie est à effacer avant envoi du projet*)* |
| --- |

**Dans le cas où vous êtes responsable de traitement et que vous en assurez vous-même les analyses (donc responsable de mise en œuvre)**, précisez-le. En effet, la réglementation (RGPD) en vigueur fait intervenir notamment deux types de rôles dans un projet d’utilisation de données de santé : le responsable de traitement est à l’initiative de l’étude ; le responsable de mise en œuvre est chargé de l’exécuter. Toutefois, en pratique, ces deux rôles peuvent être tenus par la même personne : par exemple, si vous êtes à l’initiative de l’étude (donc responsable de traitement) et que vous en assurez vous-même le traitement (donc responsable de mise en œuvre).

**En cas de co-responsabilité de traitement,** nous vous conseillons d’expliquer le partage des responsabilités dans le protocole scientifique. Ces éléments de rôles et responsabilités attenants à chaque partie seront issus d’une convention établie en amont entre ces co-responsables.

**Pour les thèses de médecine et autres mémoires**, il convient de distinguer la personne qui réalise le travail (doctorant ou doctorante) et celle qui l’encadre (directeur/trice de thèse) et qui, sous l’autorité du responsable de traitement, est le/la responsable de la mise en œuvre.

# OBJECTIFS ET FINALITÉS

## Contexte, objectif(s) et justification de l’étude

| ***Aide au remplissage***  *(*RAPPEL : cette aide au remplissage est à effacer avant envoi du projet*)* |
| --- |

Indiquer le contexte général de l’étude (scientifique, médical, organisationnel, épidémiologique, etc.). Préciser la pertinence du projet au regard de l'état actuel des connaissances et l'impact attendu.

Décrire les finalités de l’étude (amélioration de la prise en charge des patients, évaluation des politiques de santé, surveillance sanitaire…).

Indiquer le(s) objectif(s) de l’étude, les critères d’évaluation associés et la nature du(es) bénéfice(s) (scientifique, médical, social, ou concernant l’organisation du système de santé).

## Justification du respect de l’éthique

| ***Aide au remplissage***  *(*RAPPEL : cette aide au remplissage est à effacer avant envoi du projet*)* |
| --- |

Préciser en quoi l’étude est conforme aux principes éthiques, qu’elle ne pose pas de problème éthique, pas de stigmatisation d’un groupe spécifique et ne va pas à l’encontre de la morale.

## Justification de l’intérêt public

| ***Aide au remplissage***  *(*RAPPEL : cette aide au remplissage est à effacer avant envoi du projet*)* |
| --- |

Démontrer que le projet poursuit un intérêt public, notamment en indiquant :

* Les finalités du projet : elles doivent être exposées de façon claire, intelligible et sincère ;
* Le bénéfice du projet : il doit apporter un bénéfice direct ou indirect pour les individus, pour la société, ou pour la communauté scientifique ; l’intérêt commercial n’est pas incompatible avec l’intérêt public ;
* Les raisons pour lesquelles la finalité poursuivie n’entre pas dans une finalité interdite ;
* Les efforts et engagements de transparence et de publication des résultats ;
* Les dispositions prises afin de garantir l’intégrité scientifique et la qualité des études et de prévenir le risque de produire des résultats biaisés. Ces dispositions peuvent prendre la forme d’une implication d’acteurs professionnels de la recherche, de la mise en place d’une gouvernance scientifique adaptée, ou encore d’une ouverture d’éléments de méthodes ou codes (par ex. algorithmes) permettant de discuter et vérifier la validité des résultats ;
* Pour plus d’informations, vous pouvez vous reporter au site internet du Health Data Hub : [Intérêt public](https://www.health-data-hub.fr/interet-public).

| ***Définition***  *(*RAPPEL : cette partie est à effacer avant envoi du projet*)* |
| --- |

**Toutes les études, recherches et évaluations sur des données personnelles de santé doivent poursuivre une finalité d’intérêt public**. Conformément à l’article 66 de la loi Informatique et Libertés : "Les traitements relevant de la présente section ne peuvent être mis en œuvre qu'en considération de la finalité d'intérêt public qu'ils présentent. La garantie de normes élevées de qualité et de sécurité des soins de santé et des médicaments ou des dispositifs médicaux constitue une finalité d'intérêt public.” Cette exigence est rappelée au point “2.1. Finalités des traitements de la MR-004” du référentiel associé.

S’agissant des données du SNDS, en plus de poursuivre une finalité d’intérêt public, les projets doivent s’inscrire dans au moins une des finalités suivantes :

* A la recherche, aux études, à l’évaluation et à l’innovation dans les domaines de la santé et de la prise en charge médico-sociale ;
* A l’information sur la santé ainsi que sur l’offre de soins, la prise en charge médico-sociale et leur qualité;
* A la définition, à la mise en œuvre et à l’évaluation des politiques de santé et de protection sociale ;
* A la connaissance des dépenses de santé, des dépenses d’assurance maladie et des dépenses médico- sociales ;
* A l’information des professionnels, des structures et des établissements de santé ou médico-sociaux sur leur activité ;
* A la surveillance, à la veille et à la sécurité sanitaires ;

Certaines finalités sont par ailleurs strictement interdites. **Il est interdit de proposer un traitement** soit qui aurait pour objectif d’aboutir à prendre une décision concernant une personne physique identifiée sur le fondement des données la concernant et figurant l’un de ces traitements, soit qui viserait :

* La promotion en direction des professionnels de santé ou des établissements des produits de santé ;
* Ou l’exclusion de garanties des contrats d’assurance ou la modification de cotisations ou de primes d’assurance pour un individu ou un groupe d’individus.

| ***Nos conseils***  *(*RAPPEL : cette partie est à effacer avant envoi du projet*)* |
| --- |

**Si le responsable de traitement est un industriel du produit de santé ou un assureur et qu’il souhaite accéder à des données du SNDS,** il devra :

* Soit démontrer que les modalités de mise en œuvre du traitement rendent impossible toute utilisation des données pour l'une des “finalités interdites” ;
* Soit recourir à un laboratoire de recherche ou à un bureau d'études, public ou privé, pour réaliser le traitement (en tant que responsable de la mise en œuvre). Un contrat de sous-traitance reprenant les exigences de l’article 28 du RGPD devra être conclu. Ces laboratoires de recherche ou bureaux d’études réalisant des études pour des tiers devront en outre s’engager auprès de la CNIL à respecter [le référentiel](https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000035268202/) déterminant les critères de confidentialité, d’expertise et d’indépendance ([liste des organismes](https://www.health-data-hub.fr/sites/default/files/2021-05/202003-%20Engagement_de_conformite_au_reefentiel_au_06-03-20.pdf) ayant déjà fait cet engagement).

## Publication des résultats et valorisation

| ***Aide au remplissage***  *(*RAPPEL : cette aide au remplissage est à effacer avant envoi du projet*)* |
| --- |

Expliquer comment les résultats seront rendus publics : publications scientifiques ou non, communications orales (colloques, revues à comité de lecture, site Internet…).

**S’agissant des données du SNDS,** vous devez vous engager à communiquer au grand public une partie de vos résultats à l’issue de votre projet en application des obligations de transparence. Ces éléments de résultats seront rendus publics par le biais du répertoire des projets sur notre site. Pour plus d’information vous pouvez consulter le [guide](https://www.health-data-hub.fr/actualites/transparence-bonne-utilisation-des-donnees) présentant les modalités de mise en œuvre de cette obligation.

# 

# MÉTHODOLOGIE

## Design de l’étude

| ***Aide au remplissage***  *(*RAPPEL : cette aide au remplissage est à effacer avant envoi du projet*)* |
| --- |

Décrire le design de l’étude qui sera mise en place (cohorte, cas-témoins…) et indiquer s'il y aura des comparaisons le cas échéant. Préciser s'il s’agit d’analyses transversales ou longitudinales, et si l’étude est prospective ou historique.

Justifier vos choix, leur adéquation avec la question posée/l’hypothèse de recherche.

Décrire les critères qui seront étudiés en précisant quel sera le critère principal et les critères secondaires, le cas échéant.

Si une association entre une exposition et un indicateur de santé est étudiée (par exemple : association entre la consommation de tabac et apparition de maladies respiratoires), indiquer la mesure de l’association qui sera estimée (odds ratio, risque relatif…).

Les avantages du design retenu pour répondre aux objectifs de votre projet peuvent être mentionnés dans ce chapitre.

## Description et justification de la population d’étude

| ***Aide au remplissage***  *(*RAPPEL : cette aide au remplissage est à effacer avant envoi du projet*)* |
| --- |

Décrire la population de l’étude en précisant :

* Les critères d’inclusion et d’exclusion en termes de caractéristiques de la population, de pathologies visées et le périmètre géographique (régional, national, multicentrique, etc.) ;
* La période de ciblage ;
* La période d’extraction ;
* Lorsqu'un échantillonnage est effectué à partir d'une population source, il convient de fournir une description de la population source et de décrire la méthode d'échantillonnage. En cas de constitution de groupe témoin, veillez à bien préciser les critères et le ratio qui doivent être appliqués.

Justifier que les données requises sont adéquates et pertinentes au regard des finalités décrites. Les données requises sont limitées à ce qui est strictement nécessaire au regard des objectifs et finalités décrites en partie 2. - Justification de l’intérêt public : ne pas en demander plus que nécessaire. Ce principe de minimisation des données est rappelé aux points 2.2.3. et 3.2.3. Nature des données de la MR-004.

| Définition  (RAPPEL : cette partie est à effacer avant envoi du projet) |
| --- |

**Population d’étude :** population choisie en fonction de ses caractéristiques pour un projet donné.

**Cohorte :** population de sujets qui répondent à une définition donnée et qui sont suivis dans le temps.

**Extraction :** sous-ensemble de données, à un instant T, issues d’une ou des bases nécessaire(s) à l’étude (registre, cohorte, autres…).

**Période d’extraction :** période sur laquelle seront extraites les données.

**Critères de ciblage :** ensembles de règles/filtres permettant de caractériser (“cibler”) la population d’étude dont vous avez besoin pour votre étude.

**Période de ciblage :**  fenêtre temporelle sur laquelle les critères de ciblage sont appliqués pour sélectionner votre population d’étude.

| *Nos conseils*  *(RAPPEL : cette partie est à effacer avant envoi du projet)* |
| --- |

*Conseil n°1 concernant la méthode de ciblage :*

Préciser autant que possible la méthode de ciblage (algorithme d’inférence des personnes atteintes d’une certaine pathologie en absence d’un code diagnostic par exemple, combinaison de variables, etc.).

**Distinguer les données nécessaires pour l’étude elle-même (l’extraction), et celles qui peuvent être nécessaires pour identifier cette population d’étude lorsque différentes (données de ciblage).** En effet, les données de ciblage et l’extraction peuvent différer : par exemple, dans une étude visant à étudier des patients hospitalisés pour un AVC de 2012 à 2015 et à analyser leur parcours de soins jusqu’à 2019, la période de ciblage sera 2012 à 2015 et la période d’extraction, 2012 à 2019. Ces données peuvent provenir de différentes sources dès lorsqu’il n’y a pas d’appariement de données d’un même individu issues de plusieurs centres participants (cf. point 3.4).

Les critères de ciblage portent aussi bien sur les caractéristiques démographiques de la population (âge, sexe, commune de résidence, etc.) que sur des caractéristiques liées à la consommation de soins ou les hospitalisations (hospitalisation avec code CIM10, délivrance de médicament avec code CIP, etc.).

*Conseil n°2 concernant la justification de la population d’étude :*

Même déidentifiées, les données de santé restent des données personnelles. Par conséquent, leur accès doit être restreint au strict nécessaire. Nous conseillons donc de :

* Bien justifier les périodes requises, pour le ciblage, et pour l’extraction ;
* Fournir le plus de précisions possible sur la taille de la population finale : même si vous n’avez pas accès à la base de données, utilisez les données de la littérature ou des données en open source pour faire des estimations et justifier la nécessité de disposer de ce volume ;
* Décrire la représentativité de votre population d’étude ;
* Apporter une attention particulière à justifier l’utilisation de variables particulièrement sensibles (quasi-identifiantes).

*Conseil n°3 concernant la définition des critères de ciblage par sous-population :*

Aussi, pour plus de clarté, vous pouvez découper les critères de ciblage par sous-population. En cas de doute sur l'éligibilité de votre étude à une méthodologie de référence, vous pouvez consulter le [Guide pédagogique](https://health-data-hub.fr/sites/default/files/2025-10/Guide%20p%C3%A9dagogique%20Proc%C3%A9dure%20simplifi%C3%A9e%20MR-004.pptx.pdf) sur la procédure simplifiée MR-004, la [checklist MR-004](https://health-data-hub.fr/sites/default/files/2025-10/202510-%20HDH_Starterkit%20MR004%20_%20Checklist%20d%27%C3%A9valuation%20de%20la%20conformit%C3%A9%20%281%29.docx), contacter le HDH ou consulter le site de la CNIL.

## Taille de la population

## 

| Aide au remplissage  (RAPPEL : cette aide au remplissage est à effacer avant envoi du projet) |
| --- |

Fournir une estimation de la taille de la population que vous souhaitez étudier, indiquez les éléments de calculs sous-jacents.

Si vous n’avez aucune idée de la taille de la population envisagée pour l’étude, une estimation théorique est alors attendue, soit à partir d’une revue de littérature (y compris des rapports de la HAS, de Santé publique France…) soit à partir des données disponibles en open data (une section dédiée à ces données est [disponible ici](https://documentation-snds.health-data-hub.fr/open_data/)). A titre d’exemple, le site [ScanSanté](https://www.scansante.fr/) de l’ATIH permet d’estimer le nombre d’hospitalisations à partir des codes diagnostics CIM10, et les données en open data de l’Assurance Maladie permettent d’estimer le nombre de bénéficiaires ou le nombre d’exécutions par type de prestation de soins : [Open Damir,](http://open-data-assurance-maladie.ameli.fr/depenses/index.php#Open_DAMIR) [Open Bio](https://www.data.gouv.fr/fr/datasets/open-bio-base-complete-sur-les-depenses-de-biologie-medicale-interregimes/), et [Open Medic](http://open-data-assurance-maladie.ameli.fr/medicaments/index.php) notamment.

## Sources de données

## 

| Aide au remplissage  (RAPPEL : cette aide au remplissage est à effacer avant envoi du projet) |
| --- |

L’objectif de cette partie est d’expliquer à quelles données l’équipe projet souhaite accéder, pourquoi et comment. Dans le cadre de la MR-004, les données relatives aux personnes concernées (hors professionnels) par la recherche doivent provenir exclusivement :

* Des personnes elles-mêmes et/ou de leur(s) représentant(s) légal(-aux) ;
* Des professionnels intervenant dans la recherche ;
* Des bases de données légalement constituées et ayant fait l'objet des formalités nécessaires auprès des autorités compétentes

Les données de la base principale du SNDS ou de ses composantes (DCIR, PMSI, CEPIDC, SIDEP, SIVAC) ne peuvent être mobilisées pour conduire un projet dans le cadre de la MR-004 exception faite lorsque ces données sont rassemblées au sein d’un système fils conforme au référentiel de sécurité applicable au SNDS prévu par [l’arrêté du 6 mai 2024](https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000049516244) sous réserve que le projet respecte en tout point les dispositions du référentiel MR-004.

Les données des professionnels de santé intervenant dans la recherche doivent provenir exclusivement :

* des intéressés eux-mêmes ;
* de listes publiques ;
* ou de toute autre liste constituée à cette fin dans le respect des dispositions applicables.

**Pour la ou les bases de données sources,** il faudra notamment **présenter les éléments suivants** (si pertinent) :

* Le nom de la base de données source (exemple : dossiers médicaux de l’établissement X, entrepôt de l’établissement de santé X, registre de l’agence Y, etc.) ;
* Description de la base source et typologies des données disponibles ;
* Données recueillies au sein de cette ou ces bases sources ;
* Profondeur des données à analyser : période d’extraction des données nécessaires à la réalisation de l’étude avant et/ou après la période d’inclusion ;
* Si vous êtes responsable de traitement d’une de ces bases, l’autorisation CNIL associée (exemple : le registre X est autorisé par la CNIL depuis le 19 janvier 2017, autorisation n°1980120, délibération 2017-013…). Si vous n'êtes pas responsable de traitement, le nom du responsable de traitement de la base.

A l’exception des données concernant les professionnels intervenant dans la recherche, décrire la justification scientifique du traitement de l’ensemble des catégories de données. Le cas échéant, justifier le recueil et l’utilisation de données sensibles (qui pourrait impacter la vie privé en cas de réidentification des patients), à savoir les données de santé, génétiques, ou, biométriques des personnes

**Indiquez si les différentes bases de données sources ont vocation à être appariées ou non.** En cas d’appariement, préciser les variables qui seront pressenties pour le réaliser. La circulation des données pour réaliser l’appariement sera détaillée plus loin dans le protocole scientifique.

Dans le cadre de la MR-004, la recherche ne doit pas impliquer un appariement entre les données déjà existantes d'un même individu issues de plusieurs centres participants. Cependant, trois scénarios peuvent s’envisager :

* L'utilisation pour une recherche d'une base de données contenant des données issues de bases multicentriques déjà appariées antérieurement ;

Exemple : étude sur les données d’une cohorte contenant des données cliniques et des données de la base principale du SNDS déjà appariées

* L’utilisation pour une recherche de deux (ou plus) bases de données de deux (ou plus) centres participants différents dès lors qu’il n’y a pas d'appariement entre ces deux bases (car pas d’individus communs) ;

Exemple : étude sur les bases du CHU A et du CHU B contenant des dossiers médicaux sans patients communs dans les deux bases

* L'appariement dans le cadre d'une recherche entre les données d’un même individu issues d’un même centre participant.

Exemple : étude sur les données des dossiers médicaux du CHU A et les données issues d’un questionnaire soumis par le même CHU.

## Variables

## 

| Aide au remplissage  (RAPPEL : cette aide au remplissage est à effacer avant envoi du projet) |
| --- |

## Définir les variables de l’étude par type de variables (description de la population, exposition, critères, facteurs de risque, ou autres types de variables, y compris les variables de confusion et les potentiels effets modificateurs, le cas échéant). Ces catégories de données doivent correspondre à celles listées aux points 2.2.3. et 3.2.3. relatifs à la nature des données de la MR-004 et ne doivent pas en déroger.

## Préparation de données

| Aide au remplissage  (RAPPEL : cette aide au remplissage est à effacer avant envoi du projet) |
| --- |

Indiquer le ou les outils qui seront utilisés pour la préparation de données de l’étude.

Décrire les travaux de préparation des données : nettoyage, mise en qualité, redressement, annotation, dédoublonnement, corrections, etc.

Pour toutes les sources mobilisées dans le cadre de votre projet, vous devrez décrire la méthode de recueil de données, la méthode d’extraction et de préparation de ces données. Les questionnaires ou tout autre mode de recueil peuvent être décrits en annexe.

## Méthodes, traitements et analyses des données

| Aide au remplissage  (RAPPEL : cette aide au remplissage est à effacer avant envoi du projet) |
| --- |

Indiquez-le ou les outil(s) (framework de développement, software, etc.) qui seront utilisés pour analyser les données dans le cadre de l’étude ; décrire et justifier ces outils au regard de la question de recherche posée et des données disponibles.

Détailler les méthodes mathématiques ou algorithmiques (statistiques, inférence, modélisation, apprentissage machine, etc.) qui seront utilisées. Par exemple, pour un test statistique, décrire le type de test employé (Chi 2, binomial, etc.), pour les analyses multivariées (modèles utilisés, type et critère de choix des variables) ou pour un projet de machine learning la méthode d’apprentissage (SVM, Random forests, réseaux de neurones artificiels récurrents, convolutifs, etc.). Bien justifier le choix de la méthode en fonction de la nature des variables considérées, de leur volumétrie et des objectifs du projet.

Pour un projet d’apprentissage machine, décrire comment seront définis les ensembles d’apprentissage, de validation et de test, justifier leur volumétrie, indiquer si des méthodes spécifiques sont utilisées pour l’apprentissage (comme par exemple le bootstrap, le fine-tuning, …).

Si de nouvelles méthodes sont développées dans le cadre du projet ou s’il est prévu d’employer des approches qui sortent des schémas usuels, justifier pourquoi de telles approches sont envisagées et indiquer comment la qualité des résultats pourra être évaluée par comparaison avec l’état de l’art (définir des benchmarks clairs) et si nécessaire ajouter les références bibliographiques correspondantes.

Décrire les méthodes de validation et de contrôle des résultats obtenus, y compris les analyses de sensibilités, le cas échéant.

## Limites de l’étude

| Aide au remplissage  (RAPPEL : cette aide au remplissage est à effacer avant envoi du projet) |
| --- |

Identifier les risques et limites de l’étude. Détailler les biais possibles des résultats, ainsi que la méthode prévue pour les contrôler ou les limiter.

Toute limite potentielle du design de l'étude, des sources de données et des méthodes d’analyses, y compris les efforts entrepris pour réduire les biais, doit être discutée.

## Calendrier prévisionnel et faisabilité du projet

| Aide au remplissage  (RAPPEL : cette aide au remplissage est à effacer avant envoi du projet) |
| --- |

Décrire les **grandes étapes du projet et le calendrier** estimé.

Avant de définir les étapes de ce calendrier, le porteur de projet devra :

* Prendre connaissance de la délibération de la CNIL portant la procédure simplifiée MR-004 ;
* Évaluer sa conformité à cette procédure simplifiée par le remplissage de la [checklist](https://health-data-hub.fr/sites/default/files/2025-10/202510-%20HDH_Starterkit%20MR004%20_%20Checklist%20d%27%C3%A9valuation%20de%20la%20conformit%C3%A9%20%281%29.docx) dédiée ;
* Prévoir du temps de préparation des éléments à documenter pour la MR-004 (protocole, AIPD, note d’information, déclaration d’intérêt si utilisation des données du SNDS)

**Le calendrier** doit tenir compte :

* Des délais de réalisation d’un engagement de conformité sur le site de la CNIL, si ça n’a pas déjà été fait pour d’autres études ;
* De l’enregistrement de votre projet sur le répertoire public du HDH ;
* De la délivrance de l’information individuelle aux personnes ;
* Des temps de travaux de préparation des données et d’ingestion des données dans la plateforme ciblée (le cas échéant) ;
* De l’analyse des données et de l’obtention des résultats attendus ;
* Du temps nécessaire à la documentation de la base de données utilisée pour le projet, des programmes et des résultats ;
* Du temps pour valoriser les résultats obtenus, notamment sous forme de publications ou de mise en open source.

**La durée d’accès aux données doit dans ce cadre être précisée :** il s’agit de la durée nécessaire pour réaliser l’étude. Si la durée d’accès aux données est supérieure à celle nécessaire pour réaliser l’étude, préciser la date souhaitée de fin d’accès et justifier votre demande.

**Justifier la faisabilité du projet :** montrer que l’équipe dispose de l’expertise, des moyens humains, des moyens matériels et des ressources pour mener à bien le projet selon le calendrier établi ci-dessus.

| ***Nos conseils***  *(*RAPPEL : cette partie est à effacer avant envoi du projet*)* |
| --- |

Pour étayer le critère de faisabilité, nous vous conseillons de reprendre la liste des membres de l’équipe projet et pour chacun :

* décrire brièvement son apport au projet ;
* son expérience et/ou expertise en la matière (précédent projet, publications phares, formations suivies) ;
* son implication en termes de temps (% ETP).

Les compétences rassemblées dans votre équipe projet doivent couvrir toute la gamme de compétences requises.

# ANNEXES

| ***Aide au remplissage***  *(*RAPPEL : cette aide au remplissage est à effacer avant envoi du projet*)* |
| --- |

**Annexes :**

* Notice d’information individuelle soumise aux personnes concernés ; un [modèle](https://health-data-hub.fr/sites/default/files/2025-10/HDH_Starter%20Kit%20MR-004_Temp%5B%E2%80%A6%5Dion%20individuelle%20%28patients%20majeurs%29%20MR-004.docx) vous est proposé dans le cadre de la documentation pédagogique mise à votre disposition.
* Bibliographie.

**Annexe 1. Notice d’information individuelle**

**Annexe 2. Bibliographie**