Modèle de protocole scientifique

A remplir

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

Pour prendre en compte votre demande d’accès aux données de la base principale du SNDS dans le cadre de la procédure simplifiée MR-007 à accéder aux données, nous avons besoin que vous remplissiez ce protocole en respectant la structure, le format et le nombre de pages afin de faciliter l’instruction par le [CESREES](https://www.health-data-hub.fr/cesrees),. Par ailleurs, la taille de la police doit être standard selon les usages en vigueur (Calibri 10 minimum par exemple). **Le non-respect de ces consignes peut justifier la non instruction de votre dossier.**

Le CESREES est un comité indépendant qui évaluera votre dossier au regard de la finalité et la méthodologie de la recherche ; la nécessité du recours à des données de santé à caractère personnel ; la pertinence éthique ; la qualité scientifique du projet et le caractère d'intérêt public que présente le projet le cas échéant.

Par ailleurs, nous ne pouvons garantir que des pièces complémentaires à celles prévues dans la liste des pièces du dossier, que vous souhaiteriez partager en plus, soient effectivement lues.

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, vous pouvez contacter le HDH ([hdh@health-data-hub.fr](mailto:hdh@health-data-hub.fr)), la CNIL au cours de la [permanence téléphonique](https://www.cnil.fr/fr/saisir-la-cnil/contacter-la-cnil-standard-et-permanences-telephoniques) du lundi matin (Tel : 01 53 73 22 22) 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 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 les rubriques “4.2.2 - Support des données”.

Attention, aujourd’hui 1/3 des dossiers doivent être retravaillés après un passage devant le CESREES car les éléments soumis sont insuffisants ; y consacrer le temps nécessaire vous fera gagner du temps par la suite ! Les statistiques de l’accès aux données sont disponibles [ici](https://www.health-data-hub.fr/page/tableau-de-bord).

Il paraît difficile de réunir toutes les informations nécessaires au remplissage de ce dossier en moins de 2 mois. En cas de besoin, nous pouvons vous aider ici <https://www.health-data-hub.fr/contact>.

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 sur la MR-007](https://docs.google.com/presentation/d/1ujwXSAZD_qoRfUzQtyLHKa6YiDHtGBQ_/edit#slide=id.p1) pour connaître les principales exigences de cette MR et les modalités associées pour la mettre en œuvre
* Un modèle d’[information collective](https://docs.google.com/document/d/12fL58__QlK_r1fdoJ_wSxv7h7QCYGQ4St0tcjwjzZY0/edit?usp=drive_link)
* Faire appel à la [communauté](https://entraide.health-data-hub.fr/)
* Participer aux [matinales organisées](https://www.meetup.com/fr-FR/Health-Data-Hub/) en collaboration avec Ia Cnam 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]**

**Ce document ne concerne que les demandes d’accès aux données de la base principale du Système National des Données de Santé dans le cadre de la méthodologie de référence 007 (MR-007)**

**Sommaire du protocole scientifique**

[**1. PRÉSENTATION DE L’ÉQUIPE PROJET [1 PAGE MAXIMUM] 5**](#_heading=h.30j0zll)

[**2. OBJECTIFS ET FINALITÉS [2 - 4 PAGES MAXIMUM] 7**](#_heading=h.naq8wt60i932)

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

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

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

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

[**3. MÉTHODOLOGIE [3 - 5 PAGES MAXIMUM] 10**](#_heading=h.adc6t9m4vg0z)

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

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

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

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

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

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

[3.7. Méthodes, traitements et analyses des données 13](#_heading=h.3i6jdop8v0gj)

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

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

[**4. PROTECTION DE LA VIE PRIVÉE, SÉCURITÉ ET CONFIDENTIALITÉ DES DONNÉES [10 PAGES MAXIMUM]  
(Section destinée à la CNIL) 17**](#_heading=h.qa4eqzgkhjce)

[4.1. Respect des droits des personnes concernées 17](#_heading=h.l9y0egk082h4)

[4.2. Confidentialité et sécurité des données 19](#_heading=h.tks4jinhbwbg)

[4.2.1. Gestion du risque de réidentification 19](#_heading=h.janvqp1pfxzq)

[4.2.2. Support de données 20](#_heading=h.eudpuntjj3fb)

[**5. ANNEXES 21**](#_heading=h.c83bjdw1yuhg)

1. **PRÉSENTATION DE L’ÉQUIPE PROJET [1 PAGE MAXIMUM]**

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

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

* Le / les responsable(s) de traitement sont : si besoin vous référer à la définition ci-après.
* Le / les responsable(s) de la mise en œuvre sont : si besoin vous référer à la définition ci-après. Dans le cadre de la MR-007, si le responsable de mise en œuvre est un laboratoire de recherche ou un bureau d’étude (LRBE), il doit être conforme au référentiel fixé par l’[arrêté du 17 juillet 2017](https://www.legifrance.gouv.fr/loda/id/JORFTEXT000035268202). A ce titre, un engagement auprès de la CNIL doit avoir été fait par ce dernier.

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 et adresse mail.

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 la déclaration de conformité à la MR-007 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.

La MR-007 s’applique aux responsables de traitement mettant en œuvre un traitement nécessaire à la poursuite d’une mission d’intérêt public. Seuls peuvent réaliser une déclaration attestant de la conformité à la présente méthodologie de référence le ou les responsables de traitements pour lesquels la mise en oeuvre de la recherche, étude ou évaluation dans le domaine de la santé est nécessaire à l’exécution d’une mission d’intérêt public ou relève de l’exercice de l’autorité publique dont il est investi au sens de l’article 6.1.e du RGPD et répondant aux finalités mentionnées ci-dessous. À ce titre, il s’agira le plus souvent d’organismes publics. Dans le cas d’une **responsabilité conjointe** de traitement, les responsables doivent définir leurs obligations respectives conformément à l'[article 26 du RGPD](https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre4#:~:text=du%20pr%C3%A9sent%20article.-,Article%2026%20%2D%20Responsables%20conjoints%20du%20traitement,les%20responsables%20conjoints%20du%20traitement.).

**Le responsable de mise en œuvre** est chargé 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.

**Le gestionnaire du système fils** est un organisme détenant un système d’informations hébergeant ou mettant à disposition des données du SNDS dans un environnement respectant le référentiel de sécurité du SNDS, conformément aux définitions dudit référentiel.

| ***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 également sous-traitant, est chargé de l’exécuter. Toutefois, en pratique, ces deux rôles peuvent être tenus par la même personne : par exemple, une université peut être à l’initiative de l’étude (donc responsable de traitement) et assurer le traitement des données (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 et autres mémoires**, il convient de distinguer la personne qui accédera aux données dans le cadre de son travail de thèse et son organisme de rattachement qualifié le plus souvent comme responsable de traitement (université).

1. O**BJECTIFS ET FINALITÉS [2 - 4 PAGES MAXIMUM]**
   1. **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 concerne l’organisation du système de santé).

* 1. **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.

* 1. **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, recherche et évaluation sur des données personnelles de santé doivent poursuivre une finalité d’intérêt public**. C’est 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.”

Pour être conformes à la MR-007, en plus de poursuivre une finalité d’intérêt public, les projets doivent s’inscrire dans au moins une des finalités suivantes :

* Évaluation comparative de l’offre de soins : analyses spatiales, analyses stratégiques ;
* Évolution des pratiques de prise en charge, incidence de certains facteurs dans les hospitalisations, analyses temporelles ;
* Analyses comparatives des activités de soins, études de trajectoire de patients, bassin de recrutement, devenir des patients ;
* Description et analyse des pathologies et parcours de soins des patients dans les établissements de santé ;
* Analyse du territoire de santé, des groupements hospitaliers de territoire (GHT), études de collaboration entre établissements d'un périmètre défini ;
* Analyse continue d'évaluations comparatives, meilleure adaptation de l'offre de soins, optimisation, valorisation des séjours, réalisation d'indicateurs de pilotage, stratégie ;
* Travaux de modélisation, simulation, planning, logistique hospitalière, recherche opérationnelle (analyse de données dans le but d'optimiser des organisations ou de produire des éléments d'aide à la décision pour de nouvelles organisations) ;
* Ciblage des centres et/ou réalisation d'études de faisabilité pour la réalisation d'une recherche impliquant ou n'impliquant pas la personne humaine ;
* Études épidémiologiques et/ou médico-économiques dont les études pour la préparation des dossiers de discussions et réunions avec les autorités et comités compétents, ou les études à des fins de surveillance ;;

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.
  1. **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…).

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.

Dans le cadre de la MR-007, vous devez vous engager à transmettre tous les trois ans à la CNIL ainsi qu’au CESREES, un bilan synthétisant des usages des données observés durant cette période.

1. **MÉTHODOLOGIE [3 - 5 PAGES MAXIMUM]**
   1. **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. 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 principal critère 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.

* 1. **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 attendu, il convient 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.

Il convient de justifier que les données requises sont adéquates et pertinentes au regard des finalités décrites. Les données requises sont limitées à ce qui est nécessaire au regard des finalités décrites en partie 2 - Justification de l’intérêt public : ne pas en demander plus que nécessaire. En application de l’article 5, paragraphe 1, point c, du RGPD, les données traitées doivent être pertinentes, adéquates et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées (principe de minimisation des données). A cet égard, le responsable de traitement s’engage à ne traiter que les données strictement nécessaires et pertinentes au regard des objectifs de l’étude. Dès lors, chacune des catégories de données ne peut être traitée que si leur traitement est justifié dans le protocole.

**Comme votre projet mobilise des données de la base principale du SNDS, l’expression de besoin présente en annexe 2 devra être complétée dans le cadre de la soumission de votre projet.**

| 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 de la base principale du SNDS

**Période d’extraction :** période sur laquelle seront extraites les données et nécessaire à la réalisation de l’étude avant et/ou après la période d’inclusion. Dans la cadre de la MR-007, la profondeur historique maximale ne peut être supérieure à neuf ans en plus de l’année en cours, sous réserve que les données soient diffusables par la CNAM.

**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 en 2024 à étudier des patients hospitalisés pour un AVC de 2018 à 2020 et à analyser leur parcours de soins jusqu’à 2023, la période de ciblage sera 2018 à 2020 et la période d’extraction, 2018 à 2023. Ces données peuvent provenir de différentes sources.

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 médicaments avec code CIP, etc.).

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

Même pseudonymisées, les données de santé restent des données personnelles. Par conséquent, leur accès doit être restreint au strict nécessaire. Une attention particulière sera portée à la complétude de cette partie. Insuffisamment documenté, votre projet pourrait être autorisé mais pour un périmètre plus restreint que celui demandé, ce qui pourrait mettre en risque la fiabilité de l’étude. Soyez donc bien attentif à expliquer pourquoi vous avez besoin de telles données, en particulier quand c’est une donnée sensible (par exemple localisation géographique précise) ou d’une importante profondeur.

Nous conseillons donc de :

* Bien justifier les périodes requises, pour le ciblage, et pour l’extraction ;
* Décrire la représentativité de votre population d’étude le cas échéant;
* Apporter une attention particulière à justifier l’utilisation de variables particulièrement sensibles (identifiants potentiels du SNDS qui sont le mois et l’année de naissance, la date de décès, la commune de résidence, la commune de décès, la date de soin).

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

*P*our plus de clarté, vous pouvez découper les critères de ciblage par sous-population. S’agissant de la base principale du SNDS, pour remplir l’annexe 2 (relative à l’expression de besoin SNDS) vous pouvez vous appuyer sur le guide pédagogique créé à cet effet et consulter les ressources suivantes pour vous aider : le [dictionnaire des variables](https://health-data-hub.shinyapps.io/dico-snds/), [les données synthétiques au format des données réelles](https://documentation-snds.health-data-hub.fr/formation_snds/donnees_synthetiques.html), [la documentat](https://documentation-snds.health-data-hub.fr/)ion.

En cas de doute sur l'éligibilité de votre étude à une méthodologie de référence, vous pouvez contacter le HDH ou contacter la CNIL et consulter son  [site](https://www.cnil.fr/fr/traitements-declaration-conformite?field_norme_numerotation_type_value%5B0%5D=6) .

* 1. **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 ou les données en open data de l’Assurance Maladie 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/), [Data Pathologies](https://data.ameli.fr/pages/data-pathologies/) et [Open Medic](http://open-data-assurance-maladie.ameli.fr/medicaments/index.php) notamment.

* 1. **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.

**Pour décrire les données extraites de la base principale du SNDS,** il faudra notamment **présenter**

* Le nom des composantes de la base principale du SNDS mobilisées (SNIIRAM, PMSI, CépiDc, SI-DEP et Vaccin-covid dont d’ores et déjà disponibles, les données MDPH issues de la CNSA n’alimentent pas encore la base principale du SNDS) ;
* Profondeur des données à analyser ;
  1. **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).

* 1. **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.

Vous devrez décrire la méthode de préparation de ces données.

* 1. **Méthodes, traitements et analyses des données**

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

Indiquer le ou les outils (algorithmes, 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 d’ajouter les références bibliographiques correspondantes.

Décrire les méthodes de validation et de contrôle des résultats obtenus le cas échéant.

* 1. **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.

* 1. **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.

**Le calendrier** doit tenir compte :

* Des délais d’obtention de l’avis favorable avec ou sans recommandation du CESREES et d’intervention de la Cnam pour l’extraction des données et le conventionnement;
* 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 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.

Dans le cadre de la MR-007, la durée d'accès aux données dans la solution sécurisée ou la durée de conservation dans le système fils ne doivent pas excéder la durée de l’étude et ne peuvent excéder 5 ans à compter de la dernière mise à disposition effective des données.

Cette durée peut exceptionnellement être prolongée pour une durée maximale de deux ans, sur demande motivée du responsable de traitement, adressée au CESREES, qui rend alors un nouvel avis. Cette prolongation ne peut intervenir que si la durée initiale n’est pas expirée et la convention n’est pas échue. La demande au CESREES devra donc être anticipée en conséquence. Aucun archivage des données ne peut être réalisé.

**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.

Signaler si l’équipe est **formée à l’utilisation des données** visées, de la base principale du SNDS.

| ***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. Pour plus d’information sur l’offre de formation existante relative aux données de la base principale du SNDS, nous vous invitons à consulter Ia page formation du [site](https://www.health-data-hub.fr/formations).

1. **PROTECTION DE LA VIE PRIVÉE, SÉCURITÉ ET CONFIDENTIALITÉ DES DONNÉES [10 PAGES MAXIMUM]**
   1. **Respect des droits des personnes concernées**

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

Dans le cadre d’un projet utilisant des données à [caractère personnel](https://www.cnil.fr/fr/cnil-direct/question/une-donnee-caractere-personnel-cest-quoi), les principes relatifs à la protection des données s’appliquent. Pour permettre aux personnes concernées par les données du projet d’en garder la maîtrise, ces dernières ont plusieurs droits (se reporter à la partie ci-dessous Définitions pour connaître tous les [droits des personnes](https://www.health-data-hub.fr/sites/default/files/2021-01/Mes_droits.pdf) vis-à-vis des données).

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

Les droits des personnes vis-à-vis des données :

**L’information** aux personnes concernées doit être exhaustivement conforme au [Règlement Général sur la Protection des Données](https://www.cnil.fr/fr/reglement-europeen-protection-donnees) (RGPD) et à la [Loi Informatique et Libertés](https://www.cnil.fr/fr/la-loi-informatique-et-libertes) (LIL). Elle doit être individuelle, préalable à la mise en œuvre de la recherche et spécifique à chaque projet. L’obligation d’information revient au responsable de traitement. Ne pas le faire est punissable de cinq ans d'emprisonnement et de 300 000 euros d'amende (article 226-16 du code pénal).

Le **droit d‘accès** permet à toute personne concernée par les données de demander à consulter les données la concernant et à en obtenir une copie.

Le **droit de rectification** permet à toute personne concernée par les données de demander de les faire corriger si elle constate qu’elles contiennent une erreur.

Le **droit à l’effacement** permet à toute personne concernée par les données de demander à ce qu’elles soient effacées.

Le **droit d’opposition** permet à toute personne ne souhaitant pas que les données la concernant servent à l’étude, de s'opposer à cette utilisation. Cette opposition empêche toute utilisation ou conservation de données concernée

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

Tout traitement de données à caractère personnel implique, en principe, l'information individuelle et préalable des personnes concernées. Toutefois, la MR-007 admet que le responsable de traitement puisse se prévaloir d'une exception à l’information individuelle. Dans ce cas, une information collective\* doit être diffusée auprès du public afin de les informer de la mise en œuvre de la recherche. Celle-ci doit comporter l'ensemble des mentions prévues à l'article 14 du RGPD et devra, à minima, être partagée selon les modalités suivantes :

* Diffusion de la note d’information sur le site web du responsable de traitement ainsi que, le cas échéant, du laboratoire de recherche ou bureau d’études ;
* Mise en place d’un portail de transparence lorsque le responsable de traitement réalise plusieurs études à partir des données de la base principale du SNDS. Ce portail de transparence comporte une information générale sur la base principale du SNDS et une note d’information spécifique à chaque étude mise en œuvre.

D’autres modalités d’information collective peuvent également être prévues, en fonction des caractéristiques des études réalisées (réseaux sociaux, associations de patients, communiqué de presse, etc.). De plus, le responsable de traitement s'engage à respecter le principe de transparence en inscrivant son étude au sein du répertoire public du HDH.. Cependant, elle n’est pas suffisante pour être assimilée à une information collective.

* 1. **Confidentialité et sécurité des données**
     1. **Gestion du risque de réidentification**

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

Dans cette partie, il s’agit de préciser comment est mitigé le risque de ré-identification des personnes concernées (les termes marqués d’un astérisque sont définis dans la section “Définition”) :

* Les données quasi-identifiantes\* (dates exactes de soins ou de décès) sont-elle généralisées\*, et si oui, comment ?
* Combien de temps les données sont-elles conservées ? dit autrement, une fois l’objectif de traitement des données atteint, combien de temps sont-elles conservées par l’hébergeur de données (justifiez bien ce point). Vous trouverez [ici](https://www.cnil.fr/fr/les-durees-de-conservation-des-donnees), les recommandations de la CNIL sur la durée de conservation des données. La durée de conservation s’ajoute à la durée d’accès aux données.
* Dans le cadre de la MR-007, la durée d'accès dans la solution sécurisée ou la durée de conservation dans le système fils ne doit pas excéder la durée de l’étude et ne peuvent excéder 5 ans à compter de la dernière mise à disposition effective des données. Cette durée peut exceptionnellement être prolongée pour une durée maximale de deux ans, sur demande motivée du responsable de traitement, adressée au CESREES, qui rend alors un nouvel avis. Aucun archivage des données ne peut être réalisé.

Pour mémoire, le jeu de données importées dans un espace projet spécifique à une étude seront minimisés et limités aux seules données nécessaires à l’étude. Un numéro pseudonyme unique spécifique à chaque espace projet sera généré dans les mêmes conditions de pseudonymisation que celles définies par le référentiel de sécurité applicable au SNDS précité. Par exemple, un numéro pseudonyme unique pourra être généré par une fonction de hachage cryptographique résistante aux attaques par force brute ou un générateur de nombres pseudo-aléatoires cryptographiquement sûr.

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

**Quasi-identifiant :** Un quasi-identifiant, ou identifiant indirect, est une donnée qui, prise isolément, ne permet pas de réidentifier uniquement une personne, mais dont la combinaison avec d’autres quasi-identifiants le permet (e.g. code postal, date de naissance, dates d'événements précis).

**Généralisation :** Réduction de la granularité (finesse, niveau de détail) d’une variable (e.g. remplacement du code commune par le numéro du département).

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

Il est possible que vous ne soyez pas directement en mesure de répondre à ces questions, faute d’avoir accès vous-mêmes aux données.

Vous pouvez vous inspirer des réponses suivantes, qui sont à adapter spécifiquement à votre projet.

**Pour la base principale du SNDS :**

* Les données du SNDS sont pseudonymisées et ne contiennent aucune donnée directement identifiante ;

Parmi les variables quasi-identifiantes, les suivantes sont généralisées :

* La date de naissance est tronquée pour ne conserver que l’année et le mois ;
* L’adresse de résidence est généralisée au code commune.
* Les données de la base principale du SNDS ne contiennent aucune donnée liée à l’affiliation ethnique, politique, religieuse, philosophique ou syndicale, ni donnée génétique, ni donnée biométrique, ni donnée sur la vie ou l’orientation sexuelle.
  + 1. **Support de données**

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

L’objectif de cette partie est de présenter les principales mesures techniques et organisationnelles visant à assurer la sécurité de l’analyse des données.

Dans le cadre de la MR-007, les données traitées doivent **provenir exclusivement et directement de la CNAM**, seule compétente pour extraire et transmettre les données de la base principale du SNDS, dans le strict respect de l’expression de besoins.Le traitement des données doit être réalisé soit :

* sur un espace projet du portail sécurisé de la CNAM ;
* sur un espace projet, dans un environnement maîtrisé homologué (ne doit pas avoir expiré), expertisé par la Cnil dans le cadre d’un traitement (autorisation doit dater de moins de 3 ans) et conforme au référentiel de sécurité du SNDS

Par ailleurs :

* les données sont hébergées au sein de pays membres de l'espace économique européen, sans transfert possible en dehors de l’UE ;
* elles ne sont pas accessibles depuis l’extérieur de l’UE ;
* l’organisme, ou ses sous-traitants, en charge de l’hébergement est exclusivement soumis aux lois de l’UE pour les opérations d’hébergement de l’infrastructure technique de l’environnement maîtrisé, l’administration et l’exploitation associée à ce stockage.

Il convient dans cette partie de documenter où les données seront traitées et en quoi l’environnement maîtrisé respecte les conditions ci-dessus. La dernière autorisation CNIL valable délivrée dans le cadre d’un traitement dans cet environnement maîtrisé devra être transmise par le RT ou LRBE.

1. **ANNEXES**

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

**Annexes :**

* Modèles de note d’information collective (à destination des mineurs et des majeurs) ;
* Liste des variables utilisées et l’expression de besoin de la base principale du SNDS ;
* Bibliographie.