

PROTOCOL

Octobre
2021

Projet « Parcours
infarctus du
myocarde (IDM) aigu
en Ile de France » par
appariement
probabiliste entre les
données e-MUST
2014-2018 et le SNDS,
dans le Health Data
Hub

Contenu

I. INTRODUCTION	4
1. Présentation des équipes en charge du traitement	4
2. Contexte et justification du projet Parcours IDM	5
3. Finalités de Parcours IDM	6
4. Justification de l'intérêt public	7
5. Fondements juridiques	7
II. METHODOLOGIE	9
1. Données collectées	9
A. e-MUST 2014-2018	9
B. SNDS 2009-2023	11
C. Données socio-Environnementales des territoires de vie et de prise en charge	13
2. Conservation des données	13
3. Qualité des données	14
A. Les données e-MUST	14
B. Les données SNDS	14
4. Pertinence des données recueillies	14
5. Préparation, analyse et restitution des données	16
A. Préparation des données	16
B. Analyse des données	17
C. Résultats	18
D. Limites et biais	19
E. Restitution des données et valorisation des résultats	20
III. CIRCULATION DES DONNEES	22
IV. ORGANISATION ADMINISTRATIVE	25
1. Comité de pilotage de Parcours IDM	26
2. Comité scientifique	26
V. SECURITE ET CONFIDENTIALITE	27
1. Protection des droits des patients	27
A. Information des patients	27
B. Droits des patients	28
C. Obligations des sous-traitants	30
2. Support des données et sécurité	30
VI. CALENDRIER PREVISIONNEL	32
VII. ANNEXES	33

I.Introduction

1. Présentation des équipes en charge du traitement

Rôle dans le projet	Nom et fonction	Organisme de rattachement
Responsable de traitement	Mme Amélie VERDIER - Directrice générale	Agence régionale de santé d'Ile-de-France ARS-IF
Equipe projet	M. Julien MARCHAL - Co-directeur de la Direction de l'innovation, de la recherche et de la transformation numérique (DIRNOV) Dr Axelle MENU-BRANTHOMME - Responsable du Département « Traitement de l'information - Etudes » (Maitrise d'ouvrage du projet(MOA), pilotage stratégique) Mme Seak-Hy LO - Coordinatrice d'études, Département « Traitement de l'information - Etudes » (MOA, pilotage opérationnel et scientifique)	
	Mme Sandrine RIBEIRO DE SOUSSA - Juriste – Déléguée à la protection des données (DPD) M. Dominique SOULIER - Responsable de la sécurité des systèmes d'information (RSSI) adjoint	
	Dr Sophie BATAILLE - Référente « cardiologie et BPCO » - Coordinatrice des bases e-MUST et CARDIO-ARSIF (MOA, pilotage stratégique et scientifique)	
	Dr Yves LAMBERT - Président du comité scientifique e-MUST (MOA, pilotage stratégique et scientifique)	CH Versailles - SAMU 78
Sous-traitant Responsable de la mise en œuvre	Mme Naïma MEZAOUR - Directrice du GIP SESAN	GIP SESAN
Sous-traitant Equipe projet	M. Matthieu HANF - Directeur du Département « Traitement de l'information » (Maitrise d'œuvre du projet (MOE), pilotage opérationnel et scientifique) Mme Aurélie LOYEAU, M. Brice MOITEAUX et M. Daouda BA - Data scientists du Département « Traitement de l'information » (MOE, traitement opérationnel des données) Mme Nesrine BENYAHIA - Juriste – Déléguée à la protection des données (DPD) M. Rémi TILLY - Responsable de la sécurité des systèmes d'information (RSSI)	
Sous-traitant	Equipe du Health Data Hub (HDH)	HDH
Sous-traitant	INSERM - U1018/CESP (Mme Pascale TUBERT-BITTER), unité spécialisée en biostatistiques	INSERM
Partenaires	Comité scientifique de la base e-MUST - Représentants des 8 SAMU et de la BSPP - INSERM- U970/ Paris Centre de recherche Cardiovasculaire (PARCC) (Pr Xavier JOUVEN) - Comité de pilotage de la base e-MUST	
	Association de patients « Alliance du Cœur » (M. Philippe THEBAULT)	

2. Contexte et justification du projet « Parcours IDM »

Les maladies cardio-vasculaires (MCV) sont une des principales causes de morbidité avec 3,5 millions de personnes traitées pour MCV en France en 2013. Au sein des MCV, l'infarctus du myocarde (IDM) occupe une place prépondérante (environ 115 000 IDM par an en France). Le pronostic s'est amélioré au cours des 15 dernières années grâce aux progrès thérapeutiques, à la vitesse d'intervention des SAMU/SMUR¹ et BSPP² et à l'organisation de l'offre de soins (disponibilité permanente d'unités de cardiologie interventionnelle coronaire). Malgré ces améliorations, le taux de mortalité à un an post IDM aigu est de 15 %. La coordination de tous les acteurs de la prise en charge est essentielle, du dépistage précoce des facteurs de risque vasculaire (FDR) par le médecin traitant ou le cardiologue, à l'appel au SAMU, avec la prise en charge pré-hospitalière, à la prise en charge hospitalière en cardiologie, jusqu'à la prise en charge en ambulatoire par le médecin traitant et le cardiologue. Dans ce cadre, **la disponibilité d'indicateurs de pratique fiables sur l'ensemble du parcours de la maladie ischémique chronique est indispensable pour évaluer et améliorer en continu l'ensemble de la filière.**

Pour répondre à ces défis, l'Agence régionale de santé d'Ile-de-France (ARS-IF) a mis en place depuis les années 2000 une base de données permettant le recueil continu, prospectif et exhaustif d'informations médicales sur les IDM aigus (<24h) pris en charge par les SAMU/SMUR et BSPP en IDF (base de données e-MUST³). Elle permet de décrire en détail l'événement IDM aigu en pré-hospitalier, depuis le début de la douleur thoracique jusqu'à la sortie de la première hospitalisation. Cette base de données évaluant les pratiques cliniques est exhaustive (> 90%), validée, publiée et très aboutie. Cette base de données évaluant les pratiques cliniques est exhaustive (> 90%), validée, publiée et très aboutie. Cette base de données a intégré l'entrepôt de données de santé de cardiologie d'Ile-de-France (EDS de cardiologie d'IDF), autorisé par la CNIL le 26 juillet 2021 (Décision DT-2021-018) (cf. Protocole de l'EDS de cardio d'IDF).

Pour aller plus loin, il est apparu indispensable de caractériser plus globalement le parcours des IDM aigus. Le couplage des données e-MUST avec celles du SNDS constitue une opportunité unique pour disposer à la fois de données cliniques sur l'événement IDM aigu (pré-hospitalier et hospitalisation initiale) et de données détaillées et exhaustives en amont (consommations de soins et comorbidités) et en aval (consommation de soins, ré-hospitalisations, morbidité/mortalité) de l'événement IDM aigu. **En avril 2019, le présent projet « Parcours IDM » par appariement des données e-MUST 2014-2018 à celles du SNDS, a été sélectionné comme projet pilote pour participer à la construction du Health Data Hub (HDH).** L'objectif de ce projet est de mieux comprendre et expliquer le parcours de l'IDM aigu en IDF et ses déterminants, en évaluant :

- La mortalité à 3 mois, à 1 an et jusqu'à 5 ans après l'IDM aigu ainsi que les facteurs de risque associés ;
- Les consommations de soins post-hospitalières, les complications associées jusqu'à 5 ans après l'IDM aigu et les ré-hospitalisations ;
- Les FDR vasculaire et leur prise en charge et les antécédents coronaires et leur prise en charge, jusqu'à 5 ans avant l'évènement « IDM aigu ».

¹ SAMU/SMUR : Service d'Aide Médicale Urgente / Service Mobile d'Urgence et de Réanimation

² BSPP : Brigade de Sapeurs-Pompiers de Paris

³ La base de données e-MUST est décrite en Annexe 1. Description des variables e-MUST et sur le site : http://www.cardio-arsif.org/e_must.php

La base appariée e-MUST (2014-2018) – SNDS sera construite sur la plateforme sécurisée HDH et est mise à disposition des personnes habilitées sur l'espace projet. Ceux-ci pourront déployer des analyses de parcours et construire des algorithmes d'intelligence artificielle (IA). Ces algorithmes appliqués sur des volumes importants de données, permettront d'analyser entre autre les déterminants des IDM (voir protocole scientifique « Parcours IDM ») avec des logiciels de traitements (R et Python).

3. Finalités du projet « Parcours IDM »

Le projet « Parcours IDM » permet d'aborder le parcours de la maladie ischémique chronique pour des patients ayant fait un IDM aigu, en Île-de-France, de 2014 à 2018. Ce projet mobilise simultanément

- Les données e-MUST : Données cliniques pré-hospitalières et hospitalières (pendant la prise en charge de l'IDM aigu)
- Les données SNDS (hospitalières et de ville) :
 - Données post-hospitalières (jusqu'à 5 ans après l'IDM aigu) de consommation de soins, de ré-hospitalisation et de mortalité
 - Données pré-hospitalières jusqu'à 5 ans avant l'IDM aigu sur les facteurs de risques vasculaires (FDR) et leur prise en charge et les événements coronaires et leur prise en charge.

Ce projet permet d'analyser de manière objective et quantifiée le parcours de la maladie ischémique chronique pour des patients ayant fait un IDM aigu en décrivant

- La mortalité à 3 mois, à 1 an et jusqu'à 5 ans après l'IDM aigu ainsi que les facteurs de risque associés à cette mortalité
- Les consommations de soins post-hospitalières, les complications et les ré-hospitalisations, jusqu'à 5 ans après l'IDM aigu
- La consommation de soins jusqu'à 5 ans avant l'évènement aigu (IDM aigu) concernant les FDR et leur prise en charge et les antécédents coronaires et leur prise en charge.

Au-delà de la compréhension de manière objective et quantifiée de ce parcours, ces analyses permettent

- D'évaluer l'adaptation de l'offre de soins aux besoins
- D'évaluer le respect des recommandations sur la prise en charge des FDR et sur la maladie chronique ischémique (schéma thérapeutique, observance thérapeutique, schéma de surveillance, ...)
- D'évaluer la qualité des politiques de santé en se basant sur les preuves (« evidence-based »)
- De s'engager dans un processus d'amélioration des pratiques en apportant aux professionnels de santé concernés par cette maladie ischémique chronique les informations sur la réalité de la prise en charge et de standardiser les traitements et les schémas de surveillance à l'échelle régionale, en fonction des recommandations nationales et européennes
- De mettre en place un dispositif de veille sanitaire sur cette maladie chronique
- De s'engager dans des travaux scientifiques de recherche sur le parcours des maladies chroniques.

Par ailleurs, le projet « Parcours IDM » permet à l'ARS-IF de répondre aux objectifs de cardiologie fixés dans le projet régional de santé 2018-2022 (PRS2) à savoir :

- L'amélioration des parcours de soin des maladies chroniques dont la maladie chronique ischémique et l'insuffisance cardiaque chronique souvent séquellaire, permettant de réduire les complications et les ré-hospitalisations, d'améliorer la qualité de vie et de réduire la mortalité à long terme
- La réduction de l'incidence de ces maladies chroniques par l'amélioration du dépistage précoce des FDR et de leur prise en charge
- La réduction des inégalités d'accès aux soins.

Cette approche objective et quantitative du parcours de la maladie chronique ischémique à partir d'un IDM aigu pourra se décliner dans d'autres maladies chroniques, ce qui est fondamental dans un contexte de vieillissement de la population et d'augmentation permanente de la prévalence des maladies chroniques.

4. Justification de l'intérêt public

A notre connaissance, aucun registre disposant simultanément de données cliniques pré-hospitalières et hospitalières et de consommations de soins extrahospitalières précises, détaillées et exhaustives sur une large population d'IDM aigus n'existe à ce jour en France et dans le monde.

Les résultats de ce projet devraient donc déboucher directement sur une meilleure connaissance de l'IDM aigu et de son retentissement médical à court, moyen et long termes, ainsi que sur l'amélioration et la standardisation à l'échelle régionale et nationale des méthodes de suivi, de traitement et de surveillance médicale de cette pathologie. Du fait de la qualité et l'originalité des données recueillies, une valorisation tant en santé publique que sur le plan scientifique est donc à prévoir.

Comme indiqué plus haut, les données ainsi collectées et traitées permettront d'améliorer la qualité et la pertinence des actes et des pratiques de soins, d'éclairer la politique de santé et de réguler l'offre de soins en la matière et ce conformément aux missions confiées aux ARS et mentionnées à l'article L 1431-2 du code de la santé publique.

A ce titre et conformément aux dispositions de l'article 66 de la loi informatique et libertés, « Parcours IDM » présente une finalité d'intérêt public.

5. Fondements juridiques

- Compte-tenu de ce qui précède et conformément aux dispositions de l'article 66 de la loi informatique et libertés modifiée, l'étude présente une finalité d'intérêt public.
- L'étude est mise en œuvre sur le fondement de l'article 6 1.e) du règlement UE 2016-679 du Parlement européen du 27 avril 2016 (Règlement Général sur la Protection des Données ou RGPD), celle-ci étant « nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement ».

- Elle repose également sur les dispositions de l'article 9.2 i) du RGPD puisque, comme indiqué plus haut, l'étude « est nécessaire pour des motifs d'intérêt public **dans le domaine de la santé publique** [...] aux fins de garantir des normes élevées de qualité et de sécurité des soins de santé [...] ou des dispositifs médicaux [...] ».

II. Méthodologie

Le projet « Parcours IDM » est une étude épidémiologique en population, multicentrique, prospective, non contrôlée, descriptive et étiologique (facteurs de risque) se basant sur l'appariement de deux sources de données (e-MUST et SNDS). La population de l'étude est composée des 12 500 patients inclus dans e-MUST entre 2014 et 2018. L'ensemble des données collectées dans le cadre de la base e-MUST est utilisé dans ce projet. Les données du SNDS sur les consommations de soins, les hospitalisations, les Affections de Longue Durée (ALD), les arrêts de travail et les caractéristiques des décès qui concernent les patients inclus dans e-MUST entre 2014 et 2018 sont également utilisées. Pour chaque patient inclus, les données du SNDS jusqu'à 5 ans avant l'IDM aigu et jusqu'à 5 ans après l'IDM aigu sont utilisées dans ce projet. Pour les besoins du projet, l'ensemble de ces données est conservé jusqu'en 2029.

1. Données collectées

A. Base de données e-MUST 2014-2018

a) Description de la base de données e-MUST

La base de données e-MUST est développée depuis 2000 par l'ARH-IF⁴ puis l'ARS-IF, dans le cadre du Schéma Régional d'Organisation Sanitaire (SROS) de cardiologie, pour mettre en place un recueil continu, prospectif et exhaustif recensant tous les IDM évoluant depuis moins de 24h, pris en charge par les SAMU/SMUR et BSPP de la région. Les objectifs étaient alors de :

- Permettre aux professionnels de santé de mieux connaître leurs pratiques, de se comparer et de développer une démarche de qualité,
- Permettre aux tutelles d'analyser la pertinence des actes et la qualité des pratiques,
- Constituer une plateforme de discussions et d'échange, autour de données objectives, multicentriques et territoriales.

La base de données e-MUST a obtenu une autorisation par la Commission Nationale de l'Informatique et des Libertés (CNIL) sous le numéro 685983, délivrée le 25 janvier 2000 puis a intégré l'EDS de cardiologie d'IDF, autorisé par la CNIL le 26 juillet 2021 (Décision DT-2021-018).

Depuis 2000, tous les SAMU/SMUR et BSPP d'IDF (8 SAMU, 35 SMUR, 4 BSPP) participent à la constitution de la base de données e-MUST (cf. Figure 2).

Figure 1. Localisation des SAMU, SMUR et BSPP d'Île-de-France participant à la base de données e-MUST

⁴ ARH-IF : Agence Régionale d'Hospitalisation d'Ile-de-France

Comité scientifique et des centres participants. Le Comité scientifique a pour charge de proposer les conditions de tenue et d'exploitation de la base de données et d'interpréter les résultats (cf. Protocole de l'EDS de cardio d'IDF).

b) Données utilisées

Les données relatives à l'ensemble des 12 500 patients inclus dans la base de données e-MUST entre 2014 et 2018 (environ 2 500 par an) sont utilisées dans ce projet. Cette période d'inclusion large se justifie par l'obtention d'une puissance nécessaire à l'étude du rôle de certaines expositions ou événements de santé rares (exemple : rôle de l'exposition médicamenteuse sur la survenue d'IDM).

Les critères d'inclusion de ce projet « Parcours IDM » sont ceux des données e-MUST à savoir les patients qui présentent un infarctus du myocarde aigu :

- Dont le délai « début de la douleur-appel au SAMU » < 24h
- Présentant une douleur typique > à 20 min OU une autre douleur
- Avec un sus-décalage du segment ST > 0,1 mV dans les dérivations frontales (D1, D2, D3, aVL et aVF), précordiales gauches (V4 à V6) ou postérieures (V7 à V9) OU un sus-décalage du segment ST > 0,2 mV dans les dérivations précordiales droites (V1 à V3) dans au moins 2 dérivations contiguës d'un territoire coronaire OU sous-décalage du segment ST de V1 à V3 (miroir d'un sus-décalage du segment ST inféro-postérieur) OU BBG (Bloc de Branche Gauche) récent
- Non en Arrêt Cardio-Respiratoire (ACR) au moment de la prise en charge par le SMUR et avant l'ECG (ElectroCardioGramme) qualifiant
- Non-opposition du patient à la collecte et l'analyse de ses données.

Pour ce projet « Parcours IDM », l'ensemble des données 2014-2018 collectées dans le cadre de la base de données e-MUST (à l'exception des données identifiantes : NDA et numéro patient), appariées au SNDS 2009-2023, sont nécessaires pour caractériser le parcours de la maladie ischémique chronique⁵.

B. SNDS 2009-2023

a) Description du SNDS

Le Système National des Données de Santé (SNDS) constitue une avancée considérable pour analyser et améliorer la santé de la population. Géré par la Caisse Nationale de l'Assurance Maladie (CNAM) en sa qualité de responsable de traitement, le SNDS permet de chaîner à l'échelle individuelle :

- **Les données du système national d'information inter-régimes de l'Assurance Maladie (SNIIRAM)**
 - Les données de l'ensemble des prestations remboursées en ambulatoire (actes, examens paracliniques, médicaments remboursés...) pour plus de 95% des assurés en France
 - Les données du service médical de l'Assurance maladie qui comprend les informations concernant les Affections de Longue Durée (ALD), les accidents du travail, les maladies professionnelles

⁵ Cf. Annexe 1. Description des variables e-MUST.

- **Les données du programme de médicalisation des systèmes d'information (PMSI)** qui sont transmises par chaque hôpital privé ou public et mises à disposition de la CNAM pour chaînage dans le SNIIRAM et qui concernent
 - Le lieu et les dates de chaque séjour en établissement de santé
 - Les codes diagnostic codés avec la Classification Internationale des Maladies version 10 (CIM-10)
 - Le codage en Groupe Homogène de Malades, les actes, le passage dans certains secteurs de soins (réanimation, soins continus, néonatalogie, ...)
 - Ainsi qu'un score de sévérité pour les séjours en réanimation
- **Les données des décès observés sur le territoire français**, collectées par le Centre d'épidémiologie sur les causes médicales de décès (**CépiDc**) de l'INSERM et qui ont pour point de départ la certification des causes de décès effectuée par les médecins ; dans cette base sont enregistrés pour chaque décès :
 - La date de naissance de la personne décédée
 - La date de décès
 - Le lieu de décès
 - La classification CIM10 de la cause de décès.

b) Données utilisées

Par ces différentes sources, des données démographiques minimales sont également disponibles dans le SNDS :

- âge au moment des soins,
- sexe,
- commune de résidence,
- bénéfice de la Couverture Maladie Universelle,
- régime d'Assurance Maladie.

Dans le cadre de ce projet, seules les données SNDS des personnes incluses dans la base de données e-MUST entre 2014 et 2018 seront utilisées. Plus particulièrement, les grands groupes de variables suivantes seront extraits du SNDS pour étudier le parcours de soins des IDM aigus en IdF :

- Caractéristiques générales des bénéficiaires (sexe, âge, mois et année de naissance, commune de résidence, bénéfice de la Couverture Maladie Universelle, régime d'Assurance Maladie ...)
- Dates et caractéristiques des prestations remboursées en ambulatoire (actes, examens paracliniques, médicaments remboursés ...)
- Données du service médical de l'Assurance Maladie qui comprend les informations concernant les Affections de Longue Durée (ALD) et les accidents du travail
- Données des hospitalisations en Médecine-Chirurgie-Obstétrique (PMSI MCO), en hospitalisation à domicile (PMSI HAD) et en Soins de Suite et de Réadaptation (PMSI SSR) (lieu et dates de chaque séjour en établissement de santé, codes diagnostiques, actes CCAM (Classification commune des actes médicaux) ou CSARR (Catalogue spécifique des actes de rééducation et de réadaptation), passages dans certains secteurs de soins (USIC, Réanimation, SSR spécialisé en cardiologie ...), date de sortie, ...) et actes et consultations externes des établissements ex-DG (MCO, SSR)

- Données de mortalité du CépiDC (date de décès, lieu de décès, cause médicale de décès).

C. Données socio-Environnementales des territoires de vie et de prise en charge

De plus en plus d'études démontrent l'impact prépondérant du milieu de vie et de l'environnement sur les maladies cardio-vasculaires et cela à différentes échelles géographiques (pays/région/commune) ¹⁴⁻¹⁶. Dans le cadre de l'IDM, plusieurs études qui s'attachant à décrire ce lien ont déjà été réalisées dans les pays industrialisés ^{17,18}. En France, des données agrégées fiables sur les caractéristiques du milieu de vie sont maintenant disponibles en libre accès et à différentes échelles géographiques. Le croisement de ce type d'information avec celles de ce projet pourrait ainsi permettre la découverte de résultats originaux sur l'influence des expositions socio-environnementales sur la survenue des IDM.

Dans ce projet, les communes de résidence et de prise en charge seront automatiquement géocodées pour permettre la caractérisation de ces territoires grâce aux données administratives agrégées de l'INSEE. Celles-ci sont disponibles en libre accès sur leur site internet ou sur le site internet data.gouv.fr (niveaux sociodémographiques, niveau d'urbanisation des communes, période de vacances et jours fériés ...). A partir des données à la commune, plusieurs niveaux d'agrégation de données de l'INSEE seront utilisés (région, département). **Pour pouvoir réaliser ces comparaisons, une information la plus précise possible (échelle de la commune) sur le lieu du domicile ainsi que sur le lieu de prise en charge des IDM aigus est nécessaire.** Ce géocodage rendra également possible l'utilisation future des données de nature socio-environnementales produites et disponibles en libre accès par d'autres organismes d'État tels que Météo-France ou l'IGN (climat, altitude, pollution atmosphérique, ...) potentiellement utiles pour caractériser de manière systémique les déterminants de l'IDM aigu.

Recueillir les données à la commune est un bon compromis : le niveau régional et départemental sont trop larges pour appréhender des phénomènes socio-environnementaux et une granularité plus fine (code postal, adresse) ne donne pas d'information plus précise sur la personne mais ajoute un risque de réidentification de la personne.

2. Conservation des données

L'ensemble des données précédemment décrites seront nécessaires pour mieux comprendre et expliquer l'IDM aigu en Ile-de-France et avoir une vision exhaustive, de la période de vie précédant l'IDM jusqu'au décès, des caractéristiques et du parcours de soins des patients atteints d'un IDM aigu.

Pour chaque patient, les données du SNDS jusqu'à 5 ans avant et jusqu'à 5 ans après l'IDM aigu seront utilisées et conservées pendant une durée de 5 ans à partir de la dernière mise à jour de la base e-MUST/SNDS appariée. Pour pouvoir identifier des expositions anciennes susceptibles de mener à l'IDM, il sera donc nécessaire dans ce projet de recueillir des informations très en amont de l'événement aigu (jusqu'à 5 ans avant la survenue de l'IDM). De même, pour pouvoir quantifier le devenir à long terme, (parcours de soins et mortalité en particulier) un suivi prolongé des IDM est nécessaire dans ce projet. Le programme de travail étant important, il est nécessaire d'avoir assez de temps de recherche et de recul pour explorer tous les éléments décrits dans le protocole de recherche :

- Données e-MUST pour le projet « Parcours IDM » :
 - Période 2014-2018 : 1^{er} janvier 2014 au 31 décembre 2018
 - Exploitation dans le HDH jusqu'au 31 décembre 2029

- Données SNDS (SNIIRAM, PMSI et CépiDC) pour le projet « Parcours IDM »
 - Période qui fait l'objet de l'appariement : 1^{er} janvier 2009 au 31 décembre 2023
 - Exploitation dans le HDH jusqu'au 31 décembre 2029

Au-delà de la durée de conservation, les données du projet « Parcours IDM » au sein de l'espace projet seront supprimées ainsi que les copies des données respectives de chaque responsable de données.

3. Qualité des données

A. Les données e-MUST

- Les données e-MUST font l'objet d'une mise en qualité régulière, par l'équipe en charge de la base e-MUST au SESAN, en lien avec l'évolution des recommandations scientifiques nationales et européenne sur la prise en charge des IDM aigus, les résultats des audits externes et le comité scientifique e-MUST. Les résultats sont régulièrement diffusés aux professionnels de santé pour améliorer leur pratique. La base e-MUST est scientifiquement analysée et les résultats sont régulièrement publiés. Cette base de données est exhaustive, exacte et à jour.
- Les données e-MUST 2014-2018 sont envoyées une seule fois au HDH pour être appariées aux données du SNDS. La base appariée est déposée dans l'espace projet du HDH et il n'y aura pas de mise à jour ultérieure des données e-MUST dans l'espace projet.
- Les données e-MUST utilisées dans le cadre du projet « Parcours IDM » sont des données médicales et médico-administratives déjà collectées. Elles peuvent comporter des erreurs (données manquantes ou données erronées) inhérentes au processus de recueil d'informations. Ces erreurs sont toujours présentes dans la base de données issue de l'appariement entre des données e-MUST et le SNDS. Toutefois, elles sont sans impact sur les droits et libertés des personnes.

B. Les données SNDS

- La convention de transfert des données du SNDS qui sera signée entre le HDH, l'ARS-IF et la CNAM doit notamment préciser le périmètre, la fréquence de mise à jour et la sécurisation du transfert des données. La qualité des données relève de la responsabilité des producteurs de données tels que la CNAM, des établissements de santé ou des laboratoires de biologie médicale.
- Il s'agit de données médicales et médico-administratives déjà collectées. Elles peuvent comporter des erreurs dues à la nature du processus de remontées d'informations qui seront toujours présentes dans la base de données issue de l'appariement. Toutefois, les erreurs présentes dans cette base sont sans impact sur les droits et libertés des personnes.

4. Pertinence des données recueillies

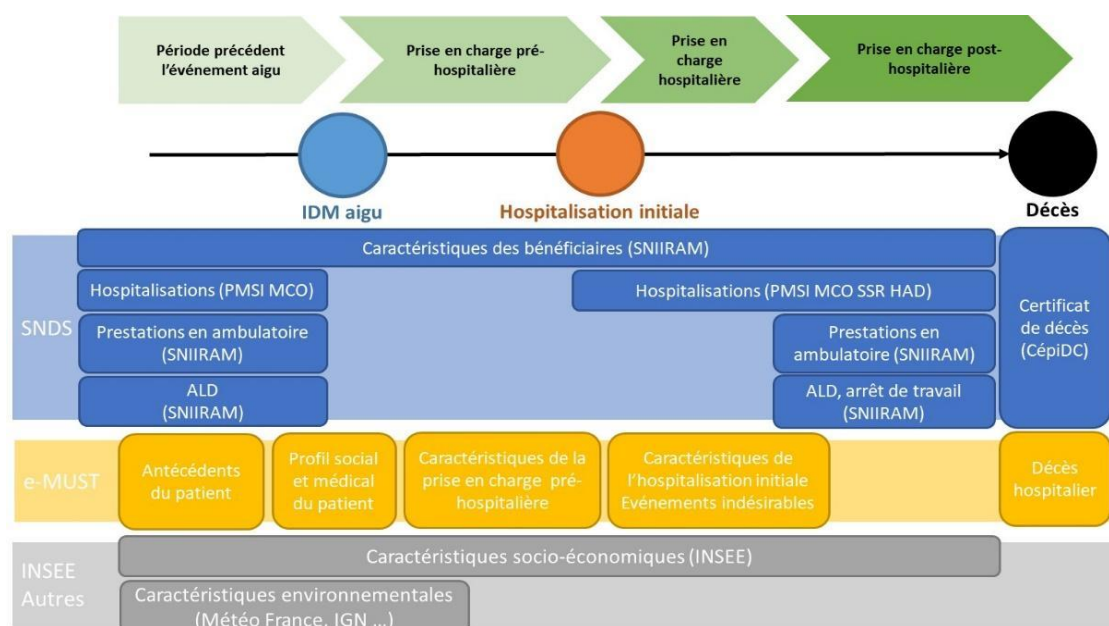
Depuis les années 2000, les données e-MUST ont alimenté les réflexions et contribué de manière objective à améliorer les pratiques des professionnels de santé impliqués dans la prise en charge des patients ayant eu un IDM aigu en Ile-de-France. Les données relatives à l'ensemble des 12 500 patients inclus dans e-MUST entre 2014 et 2018 (environ 2 500 par an) seront utilisées dans ce projet. Cette période d'inclusion large se justifie par l'obtention d'une puissance adéquate nécessaire à l'étude du rôle de certaines expositions ou événements de santé rares.

Pour ce projet, l'ensemble des données collectées dans le cadre de e-MUST seront nécessaires pour caractériser finement les parcours de soins, leurs qualité et déterminants (sauf les données identifiantes : le NDA et le numéro patient).

Pour chaque patient, les données du SNDS jusqu'à 5 ans avant et jusqu'à 5 ans après l'IDM aigu seront utilisées. En effet, depuis les travaux historiques de David Barker montrant que le risque de décès par maladie coronarienne à l'âge adulte est corrélé au poids du sujet à sa naissance, il est maintenant connu que des facteurs environnementaux et médicaux peuvent avoir un impact sur la santé à long terme se manifestant bien après l'exposition initiale. Pour pouvoir identifier des expositions anciennes susceptibles de mener à l'IDM, il sera donc nécessaire dans ce projet de recueillir des informations très en amont de l'événement aigu (jusqu'à 5 ans avant la survenue de l'IDM). De même, pour pouvoir quantifier le devenir à long terme, (parcours de soins et mortalité en particulier) un suivi prolongé des IDM est nécessaire dans ce projet (jusqu'à 5 ans après la survenue de l'IDM).

Concernant les dates de mort exactes issues du CépiDC, ces dernières sont nécessaires au jour près pour quantifier précisément la dynamique temporelle de la mortalité à court terme post-IDM. Une part non-négligeable des décès survenant souvent le jour même ou dans les jours/semaines qui suivent, la date précise de l'événement sera nécessaire pour mener des recherches pertinentes. D'autre part, il a été décrit dans la littérature scientifique, une saisonnalité mensuelle des IDM ainsi qu'une potentielle saisonnalité hebdomadaire 19. La nécessité de comprendre, suivre et analyser ce type de phénomène saisonnier (période de week-end ou de vacances par exemple), renforce la pertinence du recueil d'une information exacte (au jour près) sur les dates de décès.

Figure 2. Typologie et chronologie des données recueillies



5. Préparation, analyse et restitution des données

A. Préparation des données

Avant de commencer l'analyse, une caractérisation de la population e-MUST, non chainable aux données du SNDS sera faite pour comprendre les biais éventuels liés au chainage entre les données e-MUST et le SNDS.

Comme la population incluse dans e-MUST ne distingue pas les personnes selon leur régime d'affiliation, leur type de prise en charge, il sera nécessaire d'évaluer la compatibilité et l'exhaustivité des données comme il faut le faire au préalable pour toute analyse dans le SNDS.

Dans ce projet, la préparation des données SNDS impliquera la création d'indicateurs standards de consommation de soins à l'échelle individuelle (nombre et nature des traitements médicamenteux, de consultations générales et spécialisées, d'hospitalisations...) et cela durant différentes périodes (pré et post IDM aigu) et sur plusieurs échelles de temps (semaine/mois/années qui suivent ou précèdent l'IDM aigu). Des groupes experts incorporant des spécialistes des pathologies/événements concernés, cardiologues, médecins urgentistes et utilisateurs des données du SNIIRAM, CépiDC ou du PMSI seront constitués pour définir l'ensemble des indicateurs à compiler.

La phase de préparation des données impliquera également le développement d'algorithmes avancés de ciblage des patients incidents et prévalents dans diverses pathologies/comorbidités fréquentes, basés sur les codes diagnostiques, certains actes de ville ou hospitaliers, ainsi que l'exposition à des médicaments traceurs. Pour identifier et caractériser ces événements de santé/comorbidités à partir du SNDS, des groupes experts incorporant des spécialistes des pathologies/événements concernés, cardiologues, médecins urgentistes et utilisateurs des données du SNIIRAM, CépiDC ou du PMSI seront constitués. D'éventuelles études déjà réalisées à partir du SNDS sur les pathologies concernées seront préalablement recherchées dans la littérature. Par exemple, dans le cadre de l'identification de comorbidités telles que

les maladies endocriniennes et métaboliques, l'étude de Fosse Edorch et collaborateurs servira de point de départ au comité d'experts ²³. Dans cette étude, trois algorithmes de repérage des personnes prises en charge pour un diabète sont disponibles. Ces algorithmes sont basés soit sur les données de prise en charge pour affection de longue durée, soit sur les remboursements de traitements antidiabétiques, soit sur la combinaison de ces deux modes de prise en charge associée aux hospitalisations en lien avec le diabète. Chacun de ces algorithmes a des utilités différentes, et le choix effectué par le comité d'experts dépendra des différentes questions à l'étude. Selon les thématiques abordés, des demandes d'algorithmes spécifiques pourront également être fait auprès du réseau REDSIAM (« Réseau pour l'utilisation des données du système national des Données de Santé » : <https://www.redsiam.fr>) et de la CNAM.

B. Analyse des données

L'ensemble des analyses sera réalisé sous le logiciel libre R sur la plateforme du Health Data Hub par le SESAN, l'ARS-IF et des équipes extérieures partenaires (INSERM U970, INSERM U1018/CESP). Pour identifier les différents facteurs influençant les paramètres étudiés, des analyses descriptives et des modèles univariés et multivariés standards ayant prouvé leur robustesse seront utilisés en première ligne tout au long du projet (tels que les modèles linéaires généralisés ou le modèle de Cox par exemple).

De par le nombre et la nature hétérogène des variables étudiées, la mise en place de ce projet se confrontera à plusieurs problèmes méthodologiques non pris en compte par ces outils de première ligne. Ces problèmes comprennent entre autres l'organisation multi-niveau des données, l'hétérogénéité des populations étudiées, ainsi que les non-linéarités et interactions complexes existant entre les variables. L'utilisation d'outils statistiques adaptés aux grandes dimensions ainsi qu'aux approches causales en épidémiologie ²⁴ sera aussi primordiale pour mener à bien cette démarche.

En cas d'écart significatif à l'hypothèse de linéarité, l'utilisation de modèles additifs généralisés (GAM) ²⁵ constituera une bonne alternative à des catégorisations arbitraires de variables quantitatives ²⁶. Ces modèles ont également l'avantage de ne pas spécifier a priori la forme des relations et de modéliser finement les interactions. Des versions prenant en compte l'organisation multi-niveaux des données ²⁷ et adaptées aux grandes dimensions ²⁸ ont récemment été développées. Pour mieux caractériser l'hétérogénéité des populations étudiées, l'utilisation de méthodes telles que la régression quantile ²⁹ ou les modèles en classes latentes ³⁰ ont montré leur utilité. Des versions adaptées à l'organisation multi-niveaux des données ainsi qu'aux grandes dimensions sont également disponibles ^{31,32}. En présence d'interactions multiples, des approches de partition récursive modernes ^{33,34} pourront être utilisées. Ces dernières sont encore sous-utilisées en épidémiologie malgré leur capacité à gérer simultanément les grandes dimensions, la non-linéarité, les interactions complexes ainsi que les données manquantes ³⁵. Récemment, les réseaux bayésiens sont apparus comme des techniques alternatives aux régressions multivariées pour intégrer des données biologiques et identifier des associations statistiques complexes entre les variables ³⁶. Des problèmes de réduction de dimension vont également se poser. L'utilisation de méthodes telles que les régressions pénalisées de type Lasso ³⁷ et leurs extensions incorporant de l'information a priori sera donc nécessaire ³⁸.

Une difficulté méthodologique supplémentaire sera l'absence de groupe contrôle. Deux stratégies envisagées seront la sélection d'un groupe contrôle parmi les cas ainsi que l'utilisation de méthodes telles que l'analyse des séries de cas se basant uniquement sur les cas (Self Control Case Series, SCCS) ³⁹. Un

avantage certain des SCCS est que ces dernières ajustent implicitement sur l'ensemble des variables de confusion invariantes dans le temps. Cette méthode a été appliquée avec succès sur les données du SNDS ⁴⁰. Des versions pénalisées adaptées aux grandes dimensions ont par ailleurs été développées ⁴¹. Malheureusement, quand la mort constitue l'événement d'intérêt, une hypothèse de ces modèles est violée. Des extensions à ce type de configuration ont été mises au point ⁴². Des extensions sans définition a priori des périodes à risques ont également été proposées ⁴³.

L'utilisation du SNDS implique d'identifier les limitations éventuelles dans les inférences statistiques et d'améliorer les méthodologies d'analyse associées. Cela est particulièrement vrai concernant la modélisation des expositions médicamenteuses. Pour cela différentes stratégies sont déjà envisagées ⁴⁴–⁴⁵. Enfin, pour mieux comprendre les parcours de soins pré-per-post IDM aigu, des analyses de parcours seront aussi envisagées sur les données du SNDS (analyse séquentielle, analyse markovien, Optimal Matching Analysis...) afin de cerner les ruptures de prise en charge éventuelles ainsi que leurs déterminants ⁴⁶. Le choix de participer au HDH comme projet pilote est également motivé par les possibilités en terme de ressource (espace de calcul, logiciel R/Python, sécurité...) pour mener à bien ces analyses.

C. Résultats

L'ensemble des données précédemment décrites seront nécessaires pour 1) mieux comprendre et expliquer l'IDM aigu en IdF et 2) avoir une vision exhaustive, de la période de vie précédant l'IDM jusqu'au décès, des caractéristiques et du parcours de soins des patients atteints d'un IDM aigu. La figure 2 synthétise la typologie des données recueillies et leur chronologie.

Tableau 1. Liste non exhaustives d'indicateurs à réaliser dans le cadre du projet

	Parcours de soins pré-IDM	IDM aigu et prise en charge pré- hospitalière	Hospitalisation initiale	Parcours de soins post-hospitalier
Source de données	E-MUST /SNDS	E-MUST	E-MUST / SNDS	SNDS
Indicateurs	Type et fréquences de consultations (cardiologue/médecin généraliste, autres...) *	Fréquence des caractéristiques sociodémographiques des patients (voir descriptif données e-MUST)	Fréquence et nature des actes réalisés (coronarographie, angioplastie coronaire, échographie cardiaque, épreuve d'effort, scintigraphie, IRM, TDM...)	Fréquence et type de consultations (cardiologue/médecin généraliste, autres...) *
	Fréquence et nature des traitements médicamenteux (maladies cardiovasc., diabète, dyslipidémie, HTA, autre ...) *	Prévalence des facteurs de risque (voir descriptif données e-MUST)	Fréquence et nature des événements indésirables durant l'hospitalisation (AVC, pontage, insuffisance cardiaque, réocclusion-seconde angioplastie, chirurgie vasculaire, transfusion ...)	Fréquence et nature des traitements médicamenteux (infarctus, comorbidités, autre, ...) *

	Fréquence et nature des actes médicaux (biologie, ECG, écho. cardiaque, épreuve d'effort, IRM, TDM,...) *	Fréquence et nature des prises en charge (voir descriptif données e-MUST)	Prévalence de la mortalité hospitalière	Fréquence et nature des actes médicaux (biologie, ECG, écho. cardiaque, épreuve d'effort, IRM, TDM,...) *
	Fréquence et motifs d'hospitalisation (maladies cardiovasc., autres ...) *	Prévalence des éléments cliniques et de gravité (voir descriptif données e-MUST)	Délai des décisions et prises en charge (voir descriptif données e-MUST)	Fréquence et nature des complications (réhospitalisations, autres ...) *
	Fréquence et nature des affections longue durée (maladies cardiovasc., autres ...) *	Fréquence, nature et motifs des traitements administrés (voir descriptif données e-MUST)	Fréquence des hospitalisations liées à un IDM non incluses dans la base de données e-MUST	Fréquence et nature des affections longue durée (insuffisance cardiaque séquellaire, autres, ...) *
	Prévalence de facteurs de risques vasculaires (tabagisme, diabète, HTA, dyslipidémie, obésité ...) *	Délai des décisions et prises en charge (voir descriptif données e-MUST)	/	Prévalence de la mortalité post-hospitalière *
	Fréquence des caractéristiques sociodémographiques (dép. de résidence, couverture sociale, ...)	/	/	Fréquence des 56 groupes non exclusifs de pathologies issues de la cartographie des pathologies de la CNAM *

* ces indicateurs seront compilés à différentes échelles temporelles si nécessaire (Jours/semaines/mois/années avant IDM aigu ou en aval de l'hospitalisation initiale)

L'ensemble des données mobilisées par le projet est pseudonymisé. Les identifiants utilisés pour le projet sur la plateforme technologique du HDH sont aléatoirement générés et ne correspondent pas aux identifiants utilisés dans les bases sources, comme détaillé dans le protocole scientifique. Les exports de données sont anonymisés respectant les 3 critères d'anonymisation (non inférence, non corrélation et non individualisation)

D. Limites et biais

a) Les indicateurs construits à partir d'e-MUST

Les données e-MUST sont déclaratives et sont donc soumises à la réponse du patient. Aussi, il est possible que certains patients omettent, ne souhaitent pas ou ne soient pas en capacité de répondre à certaines questions. L'appariement aux données SNDS permettra d'objectiver certaines réponses (comorbidité, traitement, facteur de risque ...).

b) Les indicateurs construits à partir du SNDS

Les indicateurs construits à partir du SNDS n'approchent les phénomènes qu'à partir d'un remboursement de soins ou d'une déclaration faite auprès de l'assurance maladie. Ainsi, une personne ayant un remboursement d'un médicament ne signifie pas qu'il est observant et prend son traitement. De même,

les indicateurs de comorbidité via l'ALD sont relatifs puisque seuls les patients ayant fait une demande d'ALD peuvent avoir cette reconnaissance. Il est donc nécessaire de faire appel aux experts, se référer aux travaux publiés ou en cours notamment du REDSIAM publiés et de croiser plusieurs approches (ALD, consommation de médicaments...) pour cerner un phénomène qui ne pourra, de fait, prétendre être exhaustif. De même, les ALD cardiaques n'étant pas spécifiques, le croisement des approches sera d'autant plus nécessaire (croisement avec les diagnostics d'hospitalisation, de consommation de médicaments...)

Par ailleurs, certains actes ou médicaments ayant mode de financement particulier (par exemple : dotation globale) ou n'étant pas clairement identifiés dans le SNDS, certains indicateurs ne reflèteront qu'une partie de la réalité qu'il faudra bien signaler dans la méthodologie.

c) Les données contextualisant l'environnement du lieu de vie et de prise en charge du patient

Ces données caractérisent le territoire de vie et de prise en charge et non directement le patient. Le niveau socio-économique du territoire n'est donc pas la situation du patient. Cependant ces données permettent de comprendre les inégalités d'accès à certains professionnels ou d'appréhender un niveau socio-économique général qui peut expliquer certains comportements.

E. Restitution des données et valorisation des résultats

Les indicateurs (cf. liste non exhaustive dans tableau 1) ainsi que les résultats des analyses obtenus seront destinés à être restitués de manière dynamique aux acteurs de santé franciliens œuvrant dans la prise en charge des IDM aigus. Ainsi, ces professionnels pourront évaluer leur pratique de manière fluide et avec des indicateurs actualisés annuellement grâce à cette interface.

La fourniture d'une interface web dynamique permettant de consulter et comprendre facilement les implications des indicateurs/résultats obtenus sera une étape clé du projet. Ces derniers seront consultables à différentes échelles temporelles (annuelle, pluri-annuelle jusqu'à 5 années) et différents niveaux d'agrégation (région Ile-de-France et départements franciliens pour les indicateurs/résultats mis à disposition en *open data* ; établissements de santé d'Ile-de-France ou BSPP pour les résultats accessibles sous réserve d'habilitation). Tous les résultats mis à disposition seront agrégés de telle sorte qu'aucune ré-identification individuelle ne sera possible (résultat non communiqué si moins de 6 patients concernés). Un effort de développement très important sera fourni concernant la visualisation et la restitution des données pour que ces dernières soient intelligibles et permettent aux professionnels de santé d'adapter/faciliter au mieux la prise en charge des IDM. L'interface web sera intégralement réalisée en utilisant le logiciel R ou Python et la suite de package développés par Rstudio (shiny) ou Dash.

Les **articles scientifiques** seront une deuxième forme de restitution des résultats des traitements de la base « Parcours IDM aigus en Ile-de-France ».

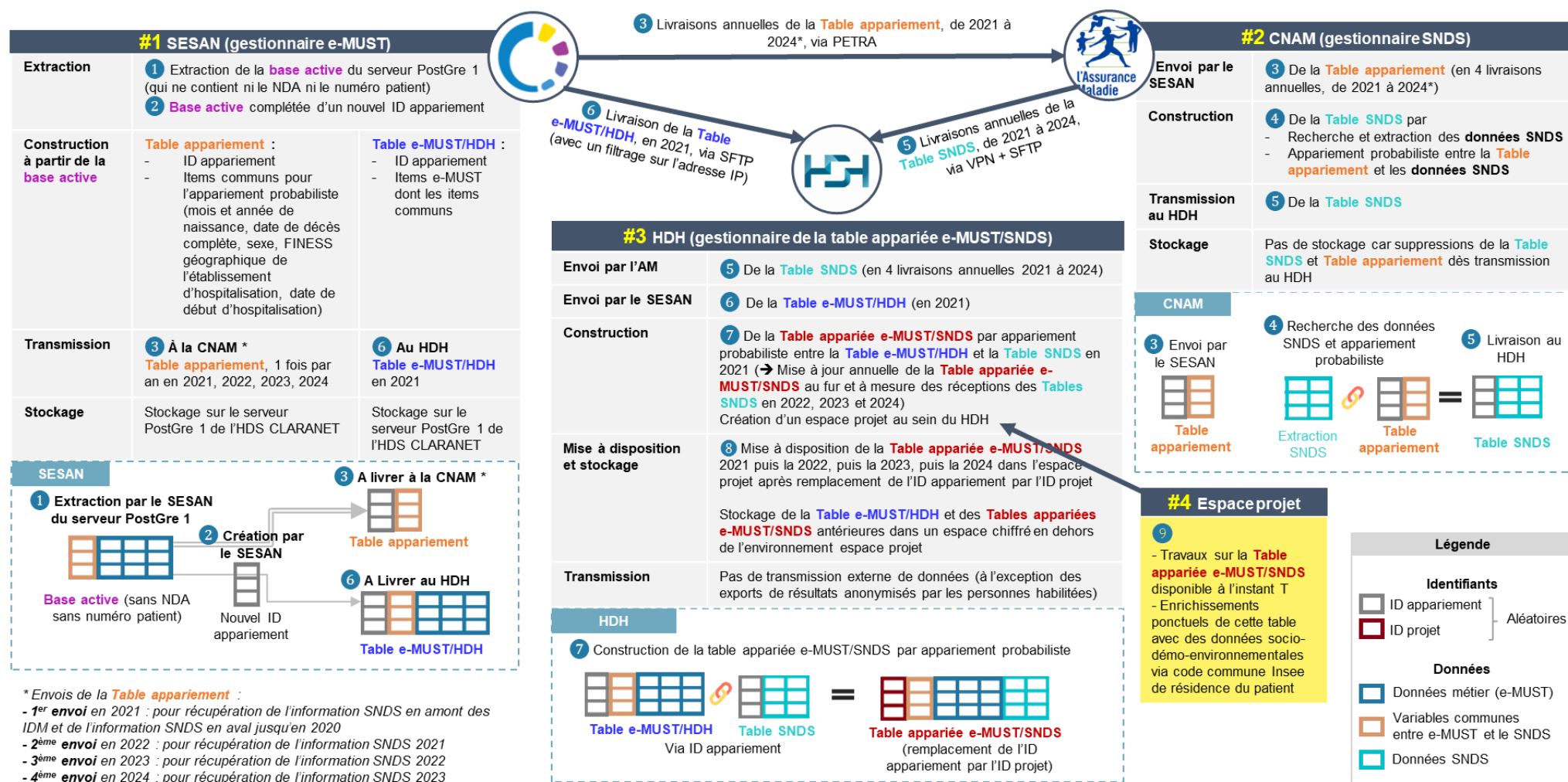
Ainsi, trois modalités de partage et de valorisation des résultats et livrables du projet sont envisagées :

- **Partage via le HdH :**
 - Algorithmes d'appariement probabiliste e-MUST & SNDS
 - Documentation technique des méthodes et outils et des données e-MUST appariées au SNDS
 - Spécifications et programmes de calcul des indicateurs du parcours ;

- **Partage via une interface web dynamique :**
 - Restitution des indicateurs, à différentes échelles géographiques et temporelles, périodiquement actualisés ;
- **Partage via publication dans des revues scientifiques :**
 - Articles scientifiques mettant en relief les résultats des traitements sur le parcours de l'IDM en Ile-de-France.

III.Circulation des données

Figure 3. Circulation et appariement des données du projet



Environnement de la circulation des données

3 espaces	2 systèmes de transfert sécurisé	5 productions
SESAN	SFTP (avec un filtrage sur l'adresse IP)	Base active
CNAM	PETRA	Table appariement
HDH		Table e-MUST/HDH
		Table SNDS
		Base appariée e-MUST/SNDS

1 Extraction de la base active (SESAN)

La base active contient l'ensemble des données e-MUST depuis 2014. Les données contenues dans cette base sont pseudonymisées (suppression des NDA et numéro de patient) et sont stockées sur le serveur PostGre 1 de l'hébergeur de données de santé (HDS) CLARANET de l'EDS de cardiologie d'IDF (Autorisation CNIL n° Décision DT-2021-018) pour limiter le risque de ré-identification en cas d'accès illégitime aux données. Une extraction de la base active est réalisée sur les années 2014-2018 dans le cadre du projet « Parcours IDM ».

2 Ajout d'un ID appariement dans la base active 2014-2018

Afin de conserver un identifiant unique pour chaque patient et de construire la table appariée e-MUST/SNDS (étape 7), un identifiant provisoire est généré dans la base active 2014-2018 de façon aléatoire et est appelé *ID appariement*. Une table de correspondance entre ID appariement et les données identifiantes des patients est conservée chiffrée au SESAN jusqu'à la dernière mise à jour de la table e-MUST/SNDS.

Deux tables sont construites à partir de la base active 2014-2018 :

- Table appariement transmise à la CNAM (étape 3) composée de
 - ID appariement et
 - items communs d'appariement probabiliste avec le SNDS (mois et année de naissance, date de décès complète, sexe, FINESS géographique de l'établissement d'hospitalisation, date de début d'hospitalisation).
- Table e-MUST/HDH transmise au Health Data Hub (HDH) (étape 6) composée de
 - ID appariement et
 - Ensemble des items e-MUST (dont les items communs).

3 Transmission de la Table appariement du SESAN vers la CNAM

Une fois par an, à partir de 2021 et jusqu'en 2024, la *Table appariement* construite, chiffrée et stockée au SESAN (étape 2) est transmise chiffrée à la CNAM via la solution PETRA (mise à disposition par la CNAM). Le premier envoi de la *Table appariement* en 2021 permet, après appariement probabiliste au SNDS (étape 4), de récupérer toute l'information disponible dans le SNDS jusqu'à 5 ans avant l'événement infarctus du myocarde (IDM) et l'information en aval jusqu'en 2020. Des deuxième, troisième et quatrième envois de la *Table appariement* sont réalisés en 2022, 2023 et 2024 et permettent de compléter la table appariée avec l'information disponible dans le SNDS de 2021, puis de 2022, puis de 2023.

NUMERO D'ENVOI	DATE D'ENVOI	RECUPERATION DES INFORMATIONS
-------------------	-----------------	-------------------------------

1	2021	L'information du parcours de soin en amont de l'événement IDM, jusqu'à 5 ans avant L'information en aval de l'événement IDM jusqu'en 2020
2	2022	Ajout de l'information 2021
3	2023	Ajout de l'information 2022
4	2024	Ajout de l'information 2023

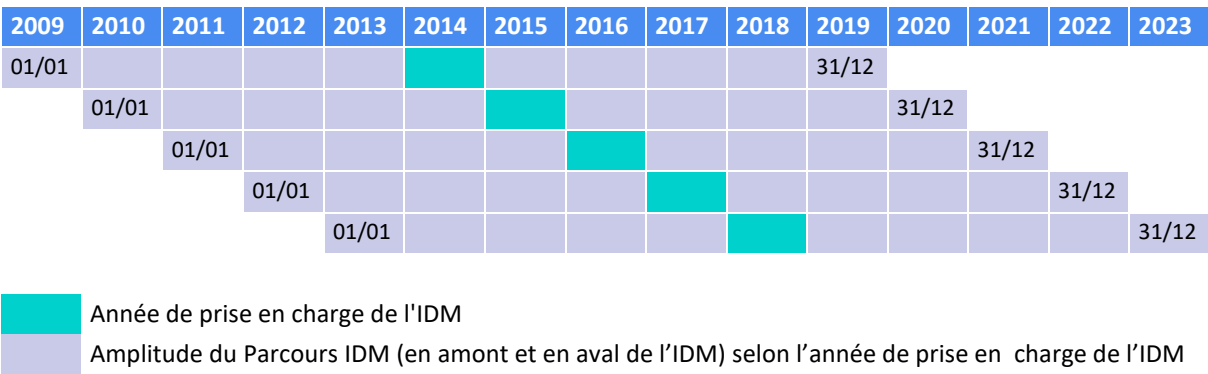
Après utilisation de la *Table appariement*, la CNAM la supprime et le SESAN la conserve chiffrée sur le PostGre 1 de l'HDS CLARANET.

4 Construction de la *Table SNDS* par appariement probabiliste entre la *Table appariement* et le SNDS (CNAM)

Après réception de la *Table appariement* par la CNAM, un appariement probabiliste entre la *Table appariement* et l'Extraction SNDS est effectué par l'équipe du DEMEX de la CNAM. Cependant, un travail collaboratif entre l'équipe du DEMEX et l'équipe ARS/SESAN est nécessaire pour choisir les meilleurs critères d'appariement et atteindre un appariement optimal.

Lorsque le taux d'appariement entre la *Table appariement* et l'extraction SNDS 2021 est jugé optimal, la CNAM fige la liste des patients de la *Table appariement* retrouvés dans le SNDS avec un identifiant CNAM puis supprime la Table SNDS et la Table appariement dès transmission au HDH. Cette liste des identifiants CNAM permet de simplifier les phases suivantes d'extraction des données du SNDS en 2022, 2023 et 2024. La CNAM extrait la *Table SNDS* composée de toutes les tables du SNDS nécessaires à la reconstitution du parcours de soins des patients e-MUST. Le « Parcours IDM » couvre une amplitude de 10 ans : 5 ans avant et après l'événement IDM.

Figure 4. Amplitude du « Parcours IDM » selon l'année de prise en charge des patients



Remarque : Seuls les patients e-MUST pris en charge en 2014 et 2015 ont leur parcours complet dès le premier envoi en 2021.

5 Transmission de la *Table SNDS* de la CNAM vers le HDH

Lorsque la *Table SNDS* (appariement probabiliste entre la *Table appariement* et l'extraction SNDS) est extraite par l'équipe du DEMEX, elle est transmise au HDH de manière **sécurisée par VPN + SFTP**, en 2021. L'opération se reproduit en 2022, puis en 2023 et en 2024.

6 Transmission de la *Table e-MUST/HDH* du SESAN vers le HDH

En 2021, le SESAN transmet, depuis le HDS Claranet, en une fois la *Table e-MUST/HDH* au HDH par SFTP (avec un filtrage sur l'adresse IP) (transmission depuis le HDS CLARANET).

Cette *Table e-MUST/HDH* permet au HDH d'accoler l'ensemble des items de la base e-MUST à la *Table SNDS* envoyée par la CNAM (étape 5) via l'*ID appariement*. Après appariement, la *Table e-MUST/HDH* est conservée chiffrée au HDH dans un environnement dédié et chiffré, séparé de l'espace projet. Elle est stockée au SESAN sur le serveur PostGre 1 de l'HDS CLARANET.

7 Construction de la *Base appariée e-MUST/SNDS* et création de l'espace projet

La *Table appariée e-MUST/SNDS* est le résultat final de la liaison entre la *Table e-MUST/HDH* (contenant l'ensemble des items e-MUST) et la *Table SNDS* (qui reconstitue avec les données du SNDS le parcours des IDM) via l'*ID appariement* (créé en étape 2). L'*ID appariement* est supprimé puis remplacé par un *ID projet* afin de ne pas faire de lien entre la base active et le SNDS en dehors de l'espace projet.

Cette *Table appariée e-MUST/SNDS* est mise à jour chaque année au fur et à mesure des reconstitutions du parcours des IDM avec la mise à jour des données du SNDS en 2022, 2023 et 2024.

8 Création de l'espace projet, mise à disposition de la Table appariée e-MUST/SNDS dans l'espace projet et stockage de la table e-MUST/HDH et des tables appariées e-MUST/SNDS antérieures

Le HDH met à disposition au porteur de projet un espace projet auquel seules les personnes habilitées au projet peuvent accéder. Le responsable de traitement (le directeur de l'ARS-IF) délègue au chef de projet « Parcours IDM » l'autorité d'habilitier les personnes faisant partie du projet et répondant aux exigences liées à l'accès à l'espace projet (ayant suivi aux sessions de sensibilisation ARS-IF et HDH, signé les CGU...). Le chef de projet vérifie l'identité de la personne, son champ d'intervention dans le projet, ses coordonnées, son email et son numéro de téléphone portable et transmet ces informations au HDH pour qu'un compte lui soit créé.

Une fois la *Table appariée e-MUST/SNDS constituée*, celle-ci est déposée sur l'espace projet « Parcours IDM » du HDH. La table e-MUST/HDH et les tables antérieures e-MUST/SNDS sont conservées chiffrées par le HDH dans un espace chiffré en dehors de l'espace projet.

9 Analyse et enrichissement de la Table appariée e-MUST/SNDS

Dans l'espace projet, seule l'équipe projet habilitée a un accès. Elle réalise :

- Des travaux sur la table appariée e-MUST/SNDS disponible à l'instant T
- Des enrichissements ponctuels de cette table avec des données socio-démo-environnementales via le code commune INSEE de résidence du patient

Cet espace est ouvert pendant toute la durée autorisée par la CNIL, jusqu'au 31 décembre 2029. Passé cette date, cet espace sera fermé et toutes les données seront supprimées.

IV. Organisation administrative

L'organisation fonctionnelle autour des données e-MUST repose sur une collaboration étroite entre les SAMU/SMUR/BSPP participants, l'ARS-IF et le SESAN. L'ARS-IF est responsable de l'EDS de cardiologie où sont collectées les données e-MUST et de sa coordination scientifique. SESAN est responsable de la gestion opérationnelle et de la valorisation des données d'e-MUST. Un Comité de pilotage et un Comité scientifique ont été créés. Le comité de pilotage est chargé de la bonne gouvernance des données et assure le suivi de la base de données (exhaustivité, qualité, fiabilité, sécurité des données) et prépare des procédures, ainsi que des rapports de synthèse à destination de l'ARS-IF, du Comité scientifique et des centres participants. Le Comité scientifique a pour charge de proposer les conditions de tenue et d'exploitation de la base de données et d'interpréter les résultats.

1. Comité de pilotage du projet « Parcours IDM »

Le comité de pilotage est composé du coordinateur de l'EDS de cardiologie d'IDF, du directeur du département Traitement de l'information – SESAN, des data scientists SESAN (responsables de la gestion et du traitement des données) et de l'attaché de recherche en sciences des données SESAN (responsable de la qualité des données)⁶.

Les principales missions du comité de pilotage sont :

- Assurer le suivi des bases de données reçues (exhaustivité, qualité et fiabilité des données)
- Analyser les rapports d'activité produits par les data scientists SESAN
- Préparer l'évolution des bases de données (développement technique, amélioration/modernisation/simplification des processus, ...)
- Vérifier que toutes les données relatives à la confidentialité sont respectées
- Entretenir un rapport étroit avec les comités scientifiques (communication, validation des décisions...).

Des réunions bi-mensuelles sont organisées entre le Health data Hub et le comité de pilotage afin de s'assurer que du bon déroulé du projet en terme de :

- Sécurité : une revue des indicateurs de sécurité et des habilitations est prévue
- De l'avancée des analyses

2. Comité scientifique e-MUST

Le comité scientifique e-MUST est composé d'un représentant de chaque SAMU, d'un représentant de la BSPP, d'un méthodologiste, d'un représentant du comité scientifique CARDIO-ARSIF et d'un représentant scientifique de l'INSERM U970. Il est représenté par le président du comité scientifique⁷.

Les principales missions du comité scientifique e-MUST sont :

- Interpréter et valider les rapports annuels d'activité présentés par le comité de pilotage
- S'assurer du respect du protocole (confidentialité des données, ...)
- Proposer les sujets scientifiques pour les articles et les communications dans les congrès
- Participer aux travaux scientifiques (rédaction des articles, communication dans les congrès)
- Proposer les évolutions de la base de données en fonction de l'évolution des pratiques et des recommandations
- Faire le lien entre les urgentistes de la région et le comité de pilotage (remontées des dysfonctionnements ou difficultés au comité de pilotage, fédération des urgentistes autour de la base de donnée).

⁶ Annexe 2. Composition des comités scientifique

⁷ Annexe 2. Composition du comité de pilotage

V.Sécurité et confidentialité

1. Protection des droits des patients

A. Information des patients

Les personnes concernées par le traitement de e-MUST sont informées du traitement des données « registre e-MUST » au moyen d'une note d'information remise lors de sa prise en charge par le SAMU/SMUR/BSPP. Elle mentionne que les données collectées servent à l'étude des pratiques des professionnels et à des travaux scientifiques⁸. Cette note d'information au patient ne mentionne pas actuellement l'appariement des données du e-MUST à celles du SNDS.

Informé *a posteriori*, les personnes prises en charge entre 2014 et 2018 du fait que leurs données collectées dans le cadre de e-MUST vont faire l'objet d'un appariement avec les données du SNDS se révélerait difficilement possible et exigerait des efforts disproportionnés car les données de e-MUST ne sont pas directement identifiantes (pas de nom, prénom, adresse de la personne prise en charge).

Ainsi, pour retrouver les personnes concernées, il serait nécessaire de faire appel à chaque centre hospitalier dans lequel la personne a été suivie suite à l'infarctus du myocarde et de demander au gestionnaire des données de retrouver une personne ayant les mêmes caractéristiques en termes de sexe, date de naissance, date d'hospitalisation, motif d'hospitalisation dans leur fichier historique. Au regard de l'activité des centres hospitaliers et du temps que ce travail demanderait pour collecter ces informations pour 12 500 personnes, une information individuelle des personnes concernées apparaît irréalisable. De même, demander aux établissements d'identifier des patients qui auraient été pris en charge par le SMUR entre 2014 et 2018 et de les informer au cours d'une nouvelle prise en charge semble impossible.

Pour ces raisons, nous demandons une dérogation à l'information individuelle des patients sur le fondement des articles 14.5b) du RGPD.

Pour pallier cette difficulté d'information individuelle, l'ARS-IF prévoit de faire afficher une note d'information sur ces traitements dans les centres de cardiologie

A toutes fins utiles, il est important de préciser qu'une publication des informations relatives au projet de recherche et aux traitements des données à caractère personnel est prévue :

- Sur le site internet présentant e-MUST (http://www.cardio-arsif.org/e_must.php)
- Sur le site internet du Health Data Hub et
- Sur le site de l'association de patients Alliance du Cœur (www.alliancecoeur.fr)

Par ailleurs, la notice d'information du traitement e-MUST délivrée aux nouveaux patients sera modifiée pour informer les personnes concernées de l'appariement des données collectées pour les données e-MUST à celles du SNDS à des fins d'études, de recherche et d'évaluation pour les collectes à venir (Annexe 2 du protocole scientifique).

⁸ Cf. Annexe 4. Note d'information individuelle au patient pour l'EDS de cardiologie d'IDF

B. Droits des patients

a) Portabilité

Le traitement des données n'étant fondé ni sur le consentement de la personne ni sur l'exécution d'un contrat, les conditions d'applicabilité du droit à la portabilité prévue par l'article 20 du RGPD ne sont pas réunies.

b) Accès

Concernant le droit d'accès aux données de l'étude, dans la mesure où celles-ci sont pseudonymisées, l'ARS-IF, par l'intermédiaire du Health Data Hub, n'est pas en mesure de remonter à l'identité des personnes et de confirmer que des données à caractère personnel les concernant sont ou ne sont pas traitées. Par conséquent, l'ARS-IF ne peut pas permettre l'accès aux dites données à caractère personnel dans le cadre du présent projet. Conformément à l'article 11 du RGPD, le Health Data Hub conserve les tables de correspondance dans le seul but de permettre aux personnes d'exercer leurs droits.

C'est la raison pour laquelle en cas d'exercice du droit d'accès par les citoyens, le Health Data Hub ou l'ARS-IF les invite à contacter le directeur de l'organisme gestionnaire du régime d'assurance maladie obligatoire auquel la personne est rattachée pour les données SNDS ou l'ARS-IF qui transmettra la demande aux établissements de santé, qui sont les seuls à disposer des identifiants directs des personnes et peuvent donc donner suite à une demande d'exercice des droits relatif aux données des bases sources. Il est également prévu que les coordonnées des correspondants à contacter pour exercer les droits soient disponibles sur le site internet du Health Data Hub et de l'ARS-IF.

c) Rectification et effacement

- Pour les mêmes raisons justifiant que l'ARS-IF n'est pas en mesure de confirmer que des données à caractère personnel d'une personne identifiée sont ou ne sont pas traitées sur la plateforme technologique, il n'a pas non plus la possibilité technique de permettre leur rectification ou leur effacement. Conformément à l'article 11 du RGPD, le Health Data Hub ne conserve pas les pseudonymes d'origine, ces derniers étant remplacés par des nouveaux identifiants projet générés aléatoirement.

Par ailleurs, l'ARS-IF n'étant pas à l'origine de la collecte des données, il ne lui est pas possible de vérifier l'exactitude des données collectées, notamment des données de santé, dans le cadre d'une demande de rectification.

- En cas d'exercice du droit de rectification ou du droit à l'effacement par les citoyens, le HDH ou l'ARS-IF les invite à contacter le directeur de l'organisme gestionnaire du régime d'assurance maladie obligatoire auquel la personne est rattachée pour les données SNDS ou l'ARS-IF qui transmettra la demande aux établissements de santé qui **sont seuls** à disposer des identifiants directs des personnes et **peuvent** donc donner suite à un exercice des droits relatif aux données des bases sources. Il est également prévu que les coordonnées des correspondants à contacter pour exercer les droits soient disponibles sur le site internet du HDH et de l'ARS-IF.
- L'exercice du droit de rectification ou du droit à l'effacement sera possible pendant la phase du projet préalable à la mise à disposition des données sur le HDH, auprès des producteurs de données sources qui sont seuls à connaître l'identité des personnes. Ainsi, en cas d'exercice du droit à l'effacement (respectivement de rectification), les producteurs de données devront supprimer de la base livrée

pour le projet de recherche les données relatives aux personnes ayant exercé leur droit à l'effacement (respectivement rectifier les données) si le producteur est en capacité de les identifier. Si le droit est exercé après la transmission des données e-MUST au SESAN, le producteur de données donne les caractéristiques d'identification du patient sur la base pseudonymisée afin que l'effacement ou la rectification soit effectué.

- Dans la mesure où des données du SNDS seront ajoutées annuellement à l'espace projet jusqu'au données de 2023, dans le cadre d'une mise à jour de type incrémental, l'exercice des droits ne s'appliquera que sur les données du SNDS nouvellement incluses dans l'espace projet de la plateforme technologique du Health Data Hub. Les données précédemment incluses ne seront pas affectées.

d) Limitation et opposition

- Pour les mêmes raisons justifiant que l'ARS-IF n'est pas en mesure de confirmer que des données à caractère personnel d'une personne identifiée sont ou ne sont pas traitées sur la plateforme technologique, il n'a pas non plus la possibilité de donner suite à une demande de limitation ou d'opposition. Conformément à l'article 11 du RGPD, le HDH ne conserve pas les pseudonymes d'origine, ces derniers étant remplacés par des nouveaux identifiants projet générés aléatoirement.
- En cas d'exercice du droit de limitation ou d'opposition par les citoyens, le HDH ou l'ARS-IF les invite à contacter le directeur de l'organisme gestionnaire du régime d'assurance maladie obligatoire auquel la personne est rattachée pour les données SNDS ou l'ARS-IF qui transmettra la demande aux établissements de santé, qui sont seuls à disposer des identifiants directs des personnes et peuvent donc donner suite à un exercice des droits relatif aux données des bases sources. Il est également prévu que les coordonnées des correspondants à contacter pour exercer les droits soient disponibles sur le site internet du HDH et de l'ARS-IF.
- L'exercice du droit de limitation ou d'opposition sera possible pendant la phase du projet préalable à la mise à disposition des données sur la plateforme technique du HDH, auprès des responsables de données qui sont seuls à connaître l'identité directe des personnes. Ainsi, en cas d'exercice du droit d'opposition, les responsables de données devront supprimer de la base source livrée pour le projet les données relatives aux personnes ayant exercé leur droit d'opposition si le responsable est en capacité de les identifier.
- Dans la mesure où des données du SNDS seront ajoutées annuellement à l'espace projet jusqu'au données de 2023, chargée en 2024, dans le cadre d'une mise à jour de type incrémental, l'exercice des droits ne s'appliquera que sur les données du SNDS nouvellement incluses dans l'espace projet de la plateforme technologique du Health Data Hub. Les données précédemment incluses ne seront pas affectées.

Dans le cadre des mises à jour incrémentales des données prévues par le projet, le HDH conserve une table d'appariement jusqu'à la dernière mise à jour. Le responsable de données fournit au fur et à mesure au HDH les identifiants temporaires d'origine correspondant à toutes les personnes ayant exercé leurs droit d'opposition (et que le responsable de données aura pu identifier dans la base d'origine). Le HDH constitue alors une matrice d'exclusion, c'est-à-dire une liste d'identifiants projet rattachés aux personnes ayant exercé leur droit d'opposition, conservée au sein de l'espace opérateur, qui permet de bloquer la mise à disposition des données concernant ces personnes lors des mises à jour.

e) En cas de transfert des données en dehors de l'Union européenne

La plateforme technologique du HDH est hébergée dans les centres de données Microsoft situés en Union Européenne, certifiés « Hébergeur de données de santé ». Compte tenu du contrat passé avec son sous-traitant et du fonctionnement des opérations d'administration de la plateforme technologique, il est possible que des données techniques d'usage de la plateforme (qui ne révèlent aucune information de santé) soient transférées vers des administrateurs situés en dehors de l'Union Européenne. Ces transferts de données sont encadrés par les clauses contractuelles types adoptées par la Commission Européenne dont une copie peut être obtenue auprès du Délégué à la protection des données du HDH.

Tableau 2. Récapitulatif d'exercice des droits de rectification, effacement, opposition des données du patient

BASE CONCERNEE	MOMENT	ACTEUR EN CHARGE DU DROIT DU PATIENT
Données e-MUST-HDH	Avant envoi « Table e-MUST-HDH » au HDH	SESAN
Données CNAM	Avant envoi « Table SNDS » au HDH	CNAM
Données e-MUST-HDH	Après appariement de « Table e-MUST-HDH » avec « Table SNDS »	Impossible
Données SNDS-HDH	Après livraison dans l'espace projet et suppression de la table de correspondance.	Impossible

C. Obligations des sous-traitants

Trois sous-traitants sont identifiés :

- Le HDH en charge de la plateforme et de l'appui technique,
- Le SESAN, en charge de la mise en œuvre du traitement
- L'INSERM qui contribue au traitement des données

Toutes ces relations sont encadrées par des conventions.

2. Support des données et sécurité

La plateforme de données de santé a reçu une homologation en janvier 2020. La plateforme du Health Data Hub respecte le référentiel de sécurité applicable au Système National des Données de Santé (SNDS) décrit dans l'arrêté du 22 mars 2017, sur les aspects à la fois organisationnels et techniques.

La démarche de sécurité pour réaliser l'homologation de la plateforme du Health Data Hub en lien avec le HFDS & l'ANSSI est en cours de réalisation. Elle couvre notamment les axes suivants :

- **Une authentification multi-facteur forte** reposant sur une infrastructure à clés publiques maîtrisée par le Ministère des Solidarités et de la Santé ;
- **Un accès à distance** à la plateforme se reposant sur l'utilisation d'une infrastructure de bureaux virtuels, garantissant la sécurité des sessions de travail ;
- **Une traçabilité** complète de l'ensemble des actions réalisées sur le Hub retranscrite dans une interface ergonomique pour les besoins d'audit ;

- **L'import et l'export sécurisé et contrôlé** de résultats, de codes et de visuels en adéquation avec les habilitations reçues.

Les données requises pour l'étude seront hébergées et traitées sur la plateforme du Health Data Hub conformément aux informations présentes dans l'AIPD de la plateforme technologique et dont la CNIL dispose.

VI. Calendrier prévisionnel

TEMPORALITE	ETAPES DU PROJET
	— Obtention de l'autorisation CNIL — Mise à disposition des données 2014-2018 de la base e-MUST pour le HDH
	— Développement d'un algorithme probabiliste puis appariement des données 2014-2018 de la base e-MUST avec les données SNDS 2009-2018 — Production de documentation des données appariées — Développement algorithmes de calcul des premiers indicateurs du parcours
	— Production des indicateurs, analyse des données (territoriale, temporelle, ...) — Documentation technique des méthodes et outils — Interface web dynamique de consultation des indicateurs
	— Appariement avec les données 2019 du SNDS — Enrichissement avec les autres données à la commune (INSEE, ...) — Poursuite du développement des algorithmes de calcul des indicateurs du parcours — Analyse des parcours pré-per-post hospitaliers — Rédaction d'un premier article scientifique
	— Appariement annuel avec les données 2020 à 2023 du SNDS — Poursuite de la valorisation scientifique des travaux — Promotion de la démarche sur d'autres bases de données (base de données francilienne des coronarographies et angioplasties, ...) — Promotion de la démarche auprès des autres régions

VII. Annexes

Annexe 1. Description des variables e-MUST.....	34
Annexe 2. Comités de pilotage et scientifiques	34
Annexe 3. Mention d'information de la collecte des données personnelles des professionnels de santé	36
Annexe 4. Note d'information individuelle au patient pour l'EDS de cardiologie d'IDF	37
Annexe 5. Mention d'information collective relative à la protection des données à caractère personnel....	38
Annexe 6. Glossaire	39
Annexe 7. Bibliographie.....	40

Annexe 1. Description des variables e-MUST

Les catégories de données e-MUST recueillies sont les suivantes¹⁴ :

- **SMUR qui prend en charge le patient (donnée de localisation)**
- Identification du patient
 - **Numéro patient e-MUST**
 - **Date de naissance complète**
 - **Sexe**
- Prise en charge
 - En primaire (ailleurs que dans un établissement de santé)
 - **Commune de prise en charge du patient (donnée de localisation)**
 - En secondaire (dans un établissement de santé)
 - **Etablissement de santé ou le patient est pris en charge initialement (donnée de localisation)**
- Antécédents (ATCD) et facteurs de risques vasculaires (FDRV)
- Sièges de l'infarctus du myocarde (IDM)
- Eléments de gravité au cours de la prise en charge pré-hospitalière
- Décisions de désobstruction prise par le SMUR
- Traitements médicamenteux en préhospitalier
- Destination du patient
 - **Etablissement de santé ou le patient est déposé (donnée de localisation)**
 - **Numéro de séjour administratif (NDA)**
- Traitements hospitaliers
- Evénements indésirables hospitaliers.

Aucune donnée non-structurée n'est recueillie dans e-MUST.

Annexe 2. Comités de pilotage et scientifiques

Composition du comité de pilotage

- Docteur Sophie BATAILLE, ARS-IF
- M. Matthieu HANF, SESAN
- Mme Aurélie LOYEAU, SESAN
- M. Brice MOITEAUX, SESAN
- Mme Fadhila IRBOUH, SESAN
- M. Daouda BA, SESAN
- Mme Seak-Hy LO, ARS-IF
- Mme Axelle MENU-BRANTHOMME

Composition du comité scientifique e-MUST

- Docteur Lionel LAMHAUT, SAMU 75
- Docteur Roger KADJI KALABANG, SAMU 77
- Docteur Yves LAMBERT, SAMU 78, Président du comité scientifique e-MUST
- Docteur Marina LAFAY, SAMU 91
- Docteur Gaëlle LEBAIL, SAMU 92
- Professeur Frédéric LAPOSTOLLE, SAMU 93

- Docteur Thévy BOCHE, SAMU 94
- Docteur François DUPAS, SAMU 95
- Docteur Jean-Michel JULIARD, représentant de la base de données CARDIO-ARSIF
- Docteur Marine SCANNAVINO, BSPP (Brigade des sapeurs-pompiers de Paris)
- Professeur Xavier JOUVEN, INSERM U970

Composition du comité scientifique CARDIO-ARSIF

- Professeur Emmanuel TEIGER, Hôpital Mondor, AP-HP
- Professeur Christian SPAULDING, HEGP, AP-HP
- Professeur Olivier VARENNE, Hôpital Cochin, AP-HP
- Docteur Christophe CAUSSIN, Institut Mutualiste Montsouris
- Docteur Bernard LIVAREK, CH de Versailles
- Docteur Gaëtan KARILLON, Hôpital Simone Veil
- Docteur François FUNCK, Hôpital René Dubos
- Docteur Philippe GAROT, ICPS Claude Galien
- Docteur Hakim BENAMER, Clinique de la Roseaie
- Docteur Yves LAMBERT, SAMU 78, représentant de la base de données e-MUST
- Professeur Xavier JOUVEN, INSERM U970

Annexe 3. Mention d'information de la collecte des données personnelles des professionnels de santé

« L'Agence régionale de santé d'Île-de-France (ARS-IF) procède à un traitement de vos données personnelles pour permettre la communication avec les professionnels de santé en charge de l'alimentation de l'Entrepôt de données santé de cardiologie d'Île-de-France (EDS de cardiologie d'IDF), la gestion des droits d'accès et assurer la sécurité du système d'information (SI).

Le traitement dans son ensemble (EDS de cardiologie d'IDF) est nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement conformément aux dispositions des articles 6.1.e) et 9.2 i) du Règlement général sur la protection des données (RGPD).

Les données à caractère personnel ont été collectées auprès de votre établissement de santé de rattachement. Les données nécessaires à la communication et à la supervision des droits d'accès au SI « EDS de cardiologie d'IDF » sont conservées le temps de votre participation au projet EDS de cardiologie d'IDF. Les données présentes dans les journaux (gestion des logs) sont conservées au maximum un an glissant par les administrateurs du SI EDS de cardiologie d'IDF.

Elles ne peuvent être communiquées qu'aux destinataires suivants : personnels habilités de l'ARS-IF et/ou de son sous-traitant SESAN.

Conformément au RGPD et à la loi n° 78-du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (loi informatique et libertés) et dans les conditions prévues par ces mêmes textes, vous disposez d'un droit d'accès et de rectification des données vous concernant, ainsi que d'un droit à demander la limitation du traitement de vos données.

Vous pouvez également vous opposer, pour des raisons tenant à votre situation particulière, au traitement des données vous concernant. Vous pouvez exercer ces droits en vous adressant au Délégué à la protection des données de l'ARS-IF par courriel à l'adresse : ars-idf-dpd@ars.sante.fr

Vous disposez d'un droit d'introduire une réclamation auprès de la Commission nationale de l'informatique et des libertés (CNIL), si vous considérez que le traitement de données à caractère personnel vous concernant constitue une violation du RGPD et de la loi informatique et libertés. »

Annexe 4. Note d'information individuelle au patient pour l'EDS de cardiologie d'IDF

Madame, Monsieur,

L'Agence Régionale de Santé Ile-de-France (ARS-IF) procède à un traitement de vos données personnelles sur le fondement des articles 6. 1. e) et 9. 2. i) du Règlement général sur la protection des données (RGPD) du 27 avril 2016.

Ce traitement a pour finalité principale, grâce à l'exploitation des données régionales, d'améliorer les pratiques de cardiologie et d'éclairer la politique régionale de santé afin de mettre en place une offre de soins adaptée. Il a également pour objectif la production d'indicateurs de cardiologie et la réalisation de travaux scientifiques de recherches et/ou d'études en santé¹ dans le respect des dispositions prévues par le RGPD et la loi n°78-du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (LIL).

Les données à caractère personnel d'identification (données pseudonymisées²) et de santé vous concernant sont collectées auprès des établissements de santé à l'occasion de votre prise en charge. Elles sont enregistrées dans un entrepôt de données de santé (EDS) sécurisé et autorisé par la CNIL et conservées pendant 20 ans à compter de la collecte et ne peuvent être communiquées qu'aux destinataires suivants :

- Personnels de l'ARS-IF et de ses sous-traitants en charge de la mise en œuvre et de la maintenance de l'EDS de cardiologie, des contrôles de qualité³, de la production des indicateurs et des statistiques et des travaux scientifiques de recherches et/ou d'études validés par le comité scientifique de l'EDS
- Professionnels de santé des établissements de santé à l'origine des données mais uniquement en ce qui concerne les données de l'établissement

Conformément au RGPD et à la LIL et dans les conditions prévues par ces mêmes textes, vous pouvez accéder aux données vous concernant ou demander leur effacement. Vous disposez également d'un droit d'opposition, d'un droit de rectification et d'un droit à la limitation du traitement de vos données.

Pour exercer ces droits, vous pouvez contacter le délégué à la protection des données de l'ARS-IF par :

- courrier à l'adresse suivante : Délégué à la protection des données de l'ARS-IF- 13, rue du Landy 93200 SAINT-DENIS, ou
- courriel à l'adresse : ars-idf-dpd@ars.sante.fr

Vous disposez enfin d'un droit d'introduire une réclamation auprès de la Commission nationale de l'informatique et des libertés (CNIL), si vous considérez que le traitement de données à caractère personnel vous concernant constitue une violation du règlement général sur la protection des données et de la loi informatique et libertés.

Paris, le 1^{er} mars 2021

¹ Pour en savoir plus et connaître la liste des travaux scientifiques de recherches et/ou d'études en santé menés à partir des données de l'EDS de cardiologie, vous pouvez consulter le site de l'ARS-IF : www.cardio-arsif.org

² Les noms, prénoms et coordonnées des patients ne sont pas collectées dans l'EDS

³ Dans le cadre des contrôles qualité, le dossier médical du patient pourra être consulté par l'attaché de recherche de sciences de la donnée (ARSD) au sein de l'établissement de santé de prise en charge, sous la responsabilité du médecin référent.

Annexe 5. Mention d'information collective relative à la protection des données à caractère personnel

Mention d'information collective relative à la protection des données à caractère personnel concernant les patients ayant eu depuis 2014 :

- un infarctus du myocarde pris en charge par les SAMU/SMUR d'IDF
- une coronarographie et/ou une angioplastie coronaire réalisée dans un des centres de cardiologie interventionnelle d'IDF

L'Agence Régionale de Santé Ile-de-France (ARS-IF) procède à un traitement de données personnelles sur le fondement des articles 6. 1. e) et 9. 2. i) du Règlement général sur la protection des données (RGPD) du 27 avril 2016.

Ce traitement a pour finalité principale, grâce à l'exploitation des données régionales, d'améliorer les pratiques de cardiologie et d'éclairer la politique régionale de santé afin de mettre en place une offre de soins adaptée. Il a également pour objectif la production d'indicateurs de cardiologie et la réalisation de travaux scientifiques de recherches et/ou d'études en santé¹ dans le respect des dispositions prévues par le RGPD et la loi n°78-du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (LIL).

Les données à caractère personnel d'identification² et de santé des patients précités sont collectées auprès des établissements de santé à l'occasion de leur prise en charge. Elles sont enregistrées dans un entrepôt de données de santé (EDS) sécurisé et autorisé par la CNIL et conservées pendant 20 ans à compter de la collecte et ne peuvent être communiquées qu'aux destinataires suivants :

- Personnels de l'ARS-IF et de ses sous-traitants en charge de la mise en œuvre et de la maintenance de l'EDS de cardiologie, des contrôles de qualité³, de la production des indicateurs et des statistiques et des travaux scientifiques de recherches et/ou d'études validés par le comité scientifique de l'EDS
- Professionnels de santé des établissements de santé à l'origine des données mais uniquement en ce qui concerne les données de l'établissement.

Conformément au RGPD et à la LIL, dans les conditions prévues par ces mêmes textes, les personnes concernées peuvent accéder à leurs données ou demander leur effacement. Elles disposent également d'un droit d'opposition, d'un droit de rectification et d'un droit à la limitation du traitement de leurs données.

Pour exercer ces droits, elles peuvent contacter le délégué à la protection des données de l'ARS-IF par :

- courrier à l'adresse suivante : Délégué à la protection des données de l'ARS-IF- 13, rue du Landy 93200 SAINT-DENIS, ou
- courriel à l'adresse : ars-idf-dpd@ars.sante.fr.

Elles disposent enfin d'un droit d'introduire une réclamation auprès de la Commission nationale de l'informatique et des libertés (CNIL), si elles considèrent que le traitement de données à caractère personnel les concernant constitue une violation du règlement général sur la protection des données et de la loi informatique et libertés.

¹ Pour en savoir plus et connaître la liste des travaux scientifiques de recherches et/ou d'études en santé menés à partir des données de l'EDS de cardiologie, vous pouvez consulter le site de l'ARS-IF : www.cardio-arsif.org

² Les noms, prénoms et coordonnées des patients ne sont pas collectées dans l'EDS

³ Dans le cadre des contrôles qualité, le dossier médical du patient pourra être consulté par l'attaché de recherche de sciences de la donnée (ARSD) au sein de l'établissement de santé de prise en charge, sous la responsabilité du médecin référent.

Annexe 6. Glossaire

Acronymes	Signification
APHP	Assistance publique - Hôpitaux de Paris
ARH-IF	Agence régionale d'hospitalisation d'Île-de-France
ARSD	Attaché de recherche en science des données
ARS-IF	Agence régionale de santé d'Île-de-France
ATCD	Antécédents
ATL	Angioplastie
BSPP	Brigade de sapeurs-pompiers de Paris
CCAM	Classification commune des actes médicaux
CEREEES	Comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé
CGU	Conditions générales d'utilisation
CHG	Centre hospitalier général
CMU	Couverture Maladie Universelle
coro	Coronarographie
CPK	Créatine PhosphoKinase
CRE	Compte rendu d'examen
CRH	Compte rendu d'hospitalisation
CSARR	Catalogue spécifique des actes de rééducation et réadaptation
DGOS	Direction générale de l'offre de soins
DPD/DPO	Délégué à la protection des données
EDS	Entrepôt de données de santé
ENRS	Environnement Numérique Régional de Santé
ESPIC	Établissement de santé privé d'intérêt collectif
FDR	Facteurs de risques
FDRV	Facteurs de risques vasculaires
GED	Gestion électronique des documents
HDS	Hébergeur de données de santé
HTA	Hyper tension artérielle
IDF	Île-de-France
IDM	Infarctus du myocarde
INSERM	Institut national de la santé et de la recherche médicale
MCV	Maladies cardio-vasculaire
NDA	Numéro de séjour du patient (numéro d'admission)
PAS	Plan d'Assurance Sécurité
PMSI	Programme de médicalisation des systèmes d'information
PRA	Plan de reprise d'activité
PRS2	Projet régional de santé 2
PSSI	Politique de sécurité des systèmes d'information
RGPD	Règlement général sur la protection des données
RSSI	Responsable de la sécurité des systèmes d'information
SAMU	Service d'aide médical urgente
SFC	Société française de cardiologie
SMUR	Service mobile d'urgence et de réanimation
SNDS	Système national des données de santé

SNIIRAM	Système national d'information inter-régimes de l'Assurance Maladie
SROS	Schéma régional de l'offre de soins
USIC	Unité de soin intensif de cardiologie
VM	Machine virtuelle
VPN	Réseau privé virtuel

Annexe 7. Bibliographie

- 1 Mendis S, Puska P, Norrving B, World Health Organization, World Heart Federation, World Stroke Organization eds. *Global atlas on cardiovascular disease prevention and control*. Geneva: World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization; 2011.
- 2 Tuppin P, Rivière S, Rigault A, Tala S, Drouin J, Pestel L, et al. Prevalence and economic burden of cardiovascular diseases in France in 2013 according to the national health insurance scheme database. *Archives of Cardiovascular Diseases* 2016;109:399–411.
- 3 Gabet A, Danchin N, Juillière Y, Olié V. Acute coronary syndrome in women: rising hospitalizations in middle-aged French women, 2004–14. *European Heart Journal* 2017;38:1060–1065.
- 4 Danchin N, Puymirat E, Aissaoui N, Adavane S, Durand E. Épidémiologie des syndromes coronaires aigus en France et en Europe. *Annales de Cardiologie et d'Angéiologie* 2010;59:S37–S41.
- 5 Lapostolle F, Bataille S, Loyeau A, Laborne F-X, Dupas F, Boche T, et al. Decision to deploy coronary reperfusion is not affected by the volume of ST-segment elevation myocardial infarction patients managed by prehospital emergency medical teams: *European Journal of Emergency Medicine* 2019;1.
- 6 Karam N, Bataille S, Marijon E, Tafflet M, Benamer H, Caussin C, et al. Incidence, Mortality, and Outcome-Predictors of Sudden Cardiac Arrest Complicating Myocardial Infarction Prior to Hospital Admission. *Circulation: Cardiovascular Interventions* 2019;12.
- 7 Loyeau A, Benamer H, Bataille S, Tepper S, Boche T, Lamhaut L, et al. Evolution of ST-Elevation Acute Myocardial Infarction Prevalence by Gender Assessed Age Pyramid Analysis—The Piramyd Study. *Journal of Clinical Medicine* 2018;7:509.
- 8 Benamer H, Bataille S, Tafflet M, Jabre P, Dupas F, Laborne FX, et al. Longer pre-hospital delays and higher mortality in women with STEMI: the e-MUST Registry. *EuroIntervention* 2016;12:e542–e549.
- 9 Karam N, Lambert Y, Tafflet M, Bataille S, Benamer H, Caussin C, et al. Changes in reperfusion strategy over time for ST segment elevation myocardial infarction in the Greater Paris Area: Results from the e-MUST Registry. *International Journal of Cardiology* 2013;168:5149–5155.
- 10 Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Revue d'Épidémiologie et de Santé Publique* 2017;65:S149–S167.
- 11 Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc J-L, Sailler L. French health insurance databases: What interest for medical research? *La Revue de Médecine Interne* 2015;36:411–417.
- 12 Villani C. *Donner un sens à l'intelligence artificielle : pour une stratégie nationale européenne*. Paris, France; 2018.
- 13 Agence Régionale de Santé Île-de-France. *Projet régional de santé 2018-2022 : améliorons la santé des franciliens*. Paris, France: ARS-IdF; 2018.
- 14 Davies AR, Grundy E, Nitsch D, Smeeth L. Constituent country inequalities in myocardial infarction incidence and case fatality in men and women in the United Kingdom, 1996-2005. *Journal of Public Health (Oxford, England)* 2011;33:131–138.
- 15 Yeh RW, Normand S-LT, Wang Y, Barr CD, Dominici F. Geographic Disparities in the Incidence and Outcomes of Hospitalized Myocardial Infarction: Does a Rising Tide Lift All Boats? *Circulation: Cardiovascular Quality and Outcomes* 2012;5:197–204.
- 16 Spatz ES, Beckman AL, Wang Y, Desai NR, Krumholz HM. Geographic Variation in Trends and Disparities in Acute Myocardial Infarction Hospitalization and Mortality by Income Levels, 1999-2013. *JAMA Cardiology* 2016;1:255.

- 17 Claey's MJ, Rajagopalan S, Nawrot TS, Brook RD. Climate and environmental triggers of acute myocardial infarction. *European Heart Journal* 2016;ehw151.
- 18 Bourdrel T, Bind M-A, Béjot Y, Morel O, Argacha J-F. Cardiovascular effects of air pollution. *Archives of Cardiovascular Diseases* 2017;110:634–642.
- 19 Manfredini R, Manfredini F, Boari B, Bergami E, Mari E, Gamberini S, et al. Seasonal and weekly patterns of hospital admissions for nonfatal and fatal myocardial infarction. *The American Journal of Emergency Medicine* 2009;27:1097–1103.
- 20 Barker DJ. Fetal origins of coronary heart disease. *BMJ (Clinical research ed.)* 1995;311:171–174.
- 21 Thornburg KL. The programming of cardiovascular disease. *Journal of Developmental Origins of Health and Disease* 2015;6:366–376.
- 22 Coloma PM, Schuemie MJ, Trifirò G, Furlong L, van Mulligen E, Bauer-Mehren A, et al. Drug-Induced Acute Myocardial Infarction: Identifying 'Prime Suspects' from Electronic Healthcare Records-Based Surveillance System. . Editor: Pant AB: *PLoS ONE* 2013;8:e72148.
- 23 Fosse-Edorh S, Rigou A, Morin S, Fezeu L, Mandereau-Bruno L, Fagot-Campagna A. Algorithmes basés sur les données médico-administratives dans le champ des maladies endocriniennes, nutritionnelles et métaboliques, et en particulier du diabète. *Revue d'Épidémiologie et de Santé Publique* 2017;65:S168–S173.
- 24 Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. *American Journal of Public Health* 2005;95:S144–S150.
- 25 Hastie T, Tibshirani R. Generalized Additive Models. *Statistical Science* 1986;1:297–310.
- 26 Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Medical Research Methodology* 2012;12.
- 27 Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1999;61:381–400.
- 28 Meier L, van de Geer S, Bühlmann P. High-dimensional additive modeling. *The Annals of Statistics* 2009;37:3779–3821.
- 29 Koenker R. *Quantile regression*. Cambridge ; New York: Cambridge University Press; 2005.
- 30 McCutcheon AL. *Latent class analysis*. Newbury Park: Sage Publications; 1987.
- 31 Geraci M, Bottai M. Linear quantile mixed models. *Statistics and Computing* 2014;24:461–479.
- 32 Belloni A, Chernozhukov V. ℓ_1 -penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics* 2011;39:82–130.
- 33 Breiman L. Random Forests. *Machine Learning* 2001;45:5–32.
- 34 Elith J, H. Graham C, P. Anderson R, Dudík M, Ferrier S, Guisan A, et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 2006;29:129–151.
- 35 Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 2009;14:323–348.
- 36 Lewis FI, Ward MP. Improving epidemiologic data analyses through multivariate regression modelling. *Emerging Themes in Epidemiology* 2013;10:4.
- 37 Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. *European Journal of Epidemiology* 2009;24:733–736.
- 38 Ahmed I, Pariente A, Tubert-Bitter P. Class-imbalanced subsampling lasso algorithm for discovering adverse drug reactions. *Statistical Methods in Medical Research* 2016.
- 39 Farrington P, Rush M, Miller E, Pugh S, Colville A, Flower A, et al. A new method for active surveillance of adverse events from diphtheria/tetanus/pertussis and measles/mumps/rubella vaccines. *The Lancet* 1995;345:567–569.
- 40 Hanf M, Quantin C, Farrington P, Benzenine E, Hocine NM, Velten M, et al. Validation of the French national health insurance information system as a tool in vaccine safety assessment: application to febrile convulsions after pediatric measles/mumps/rubella immunization. *Vaccine* 2013;31:5856–5862.
- 41 Simpson SE, Madigan D, Zorych I, Schuemie MJ, Ryan PB, Suchard MA. Multiple Self-Controlled Case Series for Large-Scale Longitudinal Observational Databases: Multiple Self-Controlled Case Series. *Biometrics* 2013;69:893–902.
- 42 Kuhnert R, Hecker H, Poethko-Müller C, Schlaud M, Vennemann M, Whitaker HJ, et al. A modified self-controlled case series method to examine association between multidose vaccinations and death. *Statistics in Medicine* 2011;30:666–677.
- 43 Morel M, Bacry E, Gaïffas S, Guilloux A, Leroy F. ConvSCCS: convolutional self-controlled case series model for lagged adverse event detection. *Biostatistics* 2019.
- 44 Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine* 2014;33:881–899.
- 45 Abrahamowicz M, Beauchamp M-E, Sylvestre M-P. Comparison of alternative models for linking drug exposure with adverse effects. *Statistics in Medicine* 2012;31:1014–1030.
- 46 Le Meur N, Gao F, Bayat S. Mining care trajectories using health administrative information systems: the use of state sequence analysis to assess disparities in prenatal care consumption. *BMC Health Services Research* 2015;15.

