

<p>Comité de sécurité de l'information Chambre sécurité sociale et santé</p>
--

CSI/CSSS/21/032

**DÉLIBÉRATION N° 21/038 DU 1<sup>ER</sup> JUIN 2021 RELATIVE AU TRAITEMENT DE DONNÉES À CARACTÈRE PERSONNEL PAR LA BANQUE CARREFOUR DE LA SÉCURITÉ SOCIALE ET VITO ET À LA COMMUNICATION DE DONNÉES À CARACTÈRE PERSONNEL RELATIVES À LA SANTÉ ISSUES DU SPF SANTÉ PUBLIQUE ET DE L'INAMI À VITO, VIA LA PLATEFORME HEALTHDATA.BE, DANS LE CADRE D'UNE ÉTUDE EPIDEMIOLOGIQUE PORTANT SUR LA SURVENANCE DE L'INFARCTUS DU MYOCARDE – EMIR STUDY**

Le Comité de sécurité de l'information, chambre sécurité sociale et santé ;

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

Vu la loi du 3 décembre 2017 *portant création de l'Autorité de protection des données*, notamment l'article 114, modifié par la loi du 25 mai 2018 ;

Vu la loi du 30 juillet 2018 *relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel* ;

Vu la loi du 5 septembre 2018 *instituant le comité de sécurité de l'information et modifiant diverses lois concernant la mise en œuvre du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE*, notamment l'article 97 ;

Vu la loi du 13 décembre 2006 *portant dispositions diverses en matière de santé*, en particulier l'article 42, § 2, 3°, modifié par la loi du 5 septembre 2018 ;

Vu la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, notamment les articles 5, 15 et 42 ;

Vu la loi du 21 août 2008 *relative à l'institution et à l'organisation de la plate-forme eHealth* ;

Vu la demande d'autorisation de Healthdata ;

Vu le rapport d'auditorat de la Plate-forme eHealth du 26 mai 2021 ;

Vu le rapport de monsieur Michel Deneyer ;

Émet, après délibération, la décision suivante, le 1<sup>er</sup> juin 2021 :

## **I. OBJET DE LA DEMANDE**

1. Par le passé, l'ancien Comité sectoriel de la sécurité sociale et de la santé s'est prononcé sur une étude épidémiologique portant sur le syndrome de Tako-Tsubo<sup>1</sup>. Cette étude a démontré un lien entre les variations de la qualité de l'air et la survenue dans les 24 heures d'un type particulier d'infarctus. Les chercheurs souhaitent à présent approfondir ces résultats au moyen d'un couplage de plusieurs bases de données contenant des données à caractère personnel relatives à la santé et de données environnementales fournies par VITO<sup>2</sup> et des coordonnées GPS du lieu de résidence et du lieu de travail des patients concernés.
2. Les personnes concernées sont les patients admis dans les hôpitaux belges pour un infarctus du myocarde de type STEMI (ST elevation myocardial infarction), NSTEMI (Non ST elevation myocardial infarction) ou Tako-Tsubo (cardiopathie de stress).

Deux cohortes distinctes de patients seront analysées :

- 1) les patients admis entre le 01/01/2012 et le 31/12/2018 soit un total de 81297 patients.
  - 2) les patients admis entre le 01/01/2019 et le 31/12/2021 (données disponibles en 2023).
3. La sélection des patients se fera au niveau du résumé hospitalier minimum (RHM) à partir des codes ICD correspondant aux pathologies étudiées<sup>3</sup>. Les codes ICD en nomenclature ICD 9 (2012,2013,2014) et leur correspondance en nomenclature ICD 10 (2016,2017,2018).
  4. Cette étude sera réalisée au moyen de données à caractère personnel relatives à la santé communiquées par la cellule technique de gestion des données du résumé hospitalier minimum de l'INAMI et de données à caractère personnel relatives à la santé issues du registre Qermid TUCO de l'INAMI. Ces données seront communiquées à la BCSS et à Healthdata via le NISS pseudonymisé du patient.

### **A) Données des résumés hospitaliers minimum (RHM)**

- 1) RHM administratives : date d'admission à l'hôpital, année de naissance du patient, âge du patient à l'admission, sexe, code postal du lieu de résidence du patient, le numéro d'agrément de l'hôpital, le type d'admission, l'indication par qui le patient a été adressé, le type de sortie, la date de sortie de l'hôpital, la durée de séjour, la destination, la

---

<sup>1</sup> Délibération n° 18/090 du 3 juillet 2018, modifiée le 4 septembre 2018, relative à la communication de données à caractère personnel relatives à la santé anonymisées par le SPF santé publique au CHU Saint-Pierre dans le cadre d'une étude épidémiologique portant sur le syndrome de Tako-Tsubo

<sup>2</sup> Vlaamse Instelling voor Technologisch Onderzoek.

<sup>3</sup> La liste des codes ICD nécessaires à l'étude sont en annexe.

catégorie du diagnostic principal (mdc), le diagnostic principal utilisé pour le groupe, le diagnostic relié au groupe, le niveau de sévérité et le risque de mortalité.

- 2) RHM diagnostic principal : STEMI, NSTEMI, Tako-Tsubo.
- 3) RHM diagnostics secondaires : diabète, insuffisance rénale, hypertension artérielle, insuffisance vasculaire périphérique, maladie coronaire, maladie cérébrovasculaire, tabagisme, dyslipidémie.
- 4) RHM procédure thérapeutique : procédure PCI, procédure CABG.

#### **B) Données issues du registre Qermid TUCO (INAMI)**

- 1) Données administratives : date d'admission, date pci, numéro d'agrément de l'hôpital, âge au moment de l'admission à l'hôpital, sexe du patient, code postal du lieu de résidence.
- 2) Indication du traitement (diagnostic principal) : indication principale, indication supplémentaire.
- 3) Facteur de risque cardiovasculaire (diagnostic secondaire) : taille, poids, BMI, diabète, insuffisance rénale (<30 ml/min), accident vasculaire cérébral, artériopathie périphérique.
- 4) Sévérité clinique : choc cardiogénique à l'admission, pontage aorto-coronaire planifié/urgent/non, état de sortie (vivant/mort), date de sortie si vivant, cause de décès.
- 5) Sévérité anatomique : maladie 1-2-3 vaisseaux, tronc commun, droite/codominauce/gauche, nombre de segments traités lors de la procédure, thrombose de sent, lésion, type de matériel, segment (PL 1-2-3-4).

#### **C) Données communiquées par la Banque Carrefour de la sécurité sociale (BCSS)**

5. La BCSS communiquera une sélection de données à caractère personnel pseudonymisées issues du datawarehouse marché du travail et protection sociale ainsi que les coordonnées GPS de l'adresse du lieu de résidence principal et l'adresse professionnelle du patient. Cette conversion sera réalisée par VITO en sous-traitance de la BCSS au sein des locaux de la BCSS.

- 1) Adresse de résidence principale au moment de l'infarctus géocodée en données GPS.
- 2) Adresse professionnelle au moment de l'infarctus géocodée en données GPS.
- 3) Données socioéconomiques issues du datawarehouse marché du travail et protection sociale : statut marital, catégorie socio-professionnelle (employés salariés, ouvriers salariés, fonctionnaire, indépendants, retraités, chômeurs), catégorie de revenu annuel (catégorie 1 : 1er décile, catégorie 2 : 2ème décile, catégorie 3 : 3ème décile, catégorie 4 : 4ème décile, catégorie 5 : 5ème décile, catégorie 6 : 6ème décile, catégorie 7 : 7ème décile, catégorie 8 : 8ème décile, catégorie 9 : 9ème décile catégorie 10: 10ème décile)
- 4) Statut vivant ou décédé, année du décès.

#### **D) Données issues de la base de données BIWAC**

6. L'étude nécessite également la communication de données à caractère personnel anonymisées détenues par le Belgian Interdisciplinary Working Group on Acute Cardiology (BIWAC).

1) BIWAC administratives : date d'admission, numéro d'agrément de l'hôpital, sexe, âge à l'admission

2) BIWAC indication du traitement (diagnostic principal) : type d'infarctus STEMI (antérieur, non antérieur)

3) BIWAC facteurs de risque cardiovasculaire (diagnostic secondaire) : tabagisme, diabète, hypertension artérielle, dyslipidémie, insuffisance rénale, maladie cérébrovasculaire, poids et taille

4) BIWAC sévérité clinique : score Killip (de 1 à 4), choc cardiogénique ou arrêt cardiaque à l'admission, heure du début des douleurs, heure de la reperfusion (stent ou thrombolyse), décès durant le séjour hospitalier

5) BIWAC sévérité anatomique : PCI date, PCI type (culprit vessel)

### **E) Données anonymes issues du registre ILI**

7. Des données anonymisées issues du registre ILI (influenza-like illness) seront communiquées par Sciensano.

Il s'agira de 3 types de données couvrant l'incidence des infections respiratoires à Influenza (grippe saisonnière) :

- incidence hebdomadaire des consultations pour syndromes grippaux par catégorie d'âge et par région ;
- indice de positivité pour l'influenza par catégorie d'âge et par région ;
- taux journalier de tests sérologiques positifs dans le réseau de laboratoire couvrant environ 60% des laboratoires de microbiologie Belge.

### **F) Données environnementales - VITO**

VITO fournira 6 groupes de données environnementales aux géocodes correspondant aux adresses de résidence et aux adresses professionnelles fournies par la BCSS :

1) qualité de l'air : moyenne journalière de pollution particulaire (PM 10, PM2.5), pollution gazeuse (NO<sub>2</sub>, O<sub>3</sub>) à une résolution de 4km<sup>2</sup> ;

2) température de l'air : moyenne journalière à une résolution de 9km<sup>2</sup> ;

3) exposition au bruit des axes routiers et aéroports : moyenne annuelle à la résolution GPS ;

4) index de végétalisation : moyenne annuelle à la résolution GPS ;

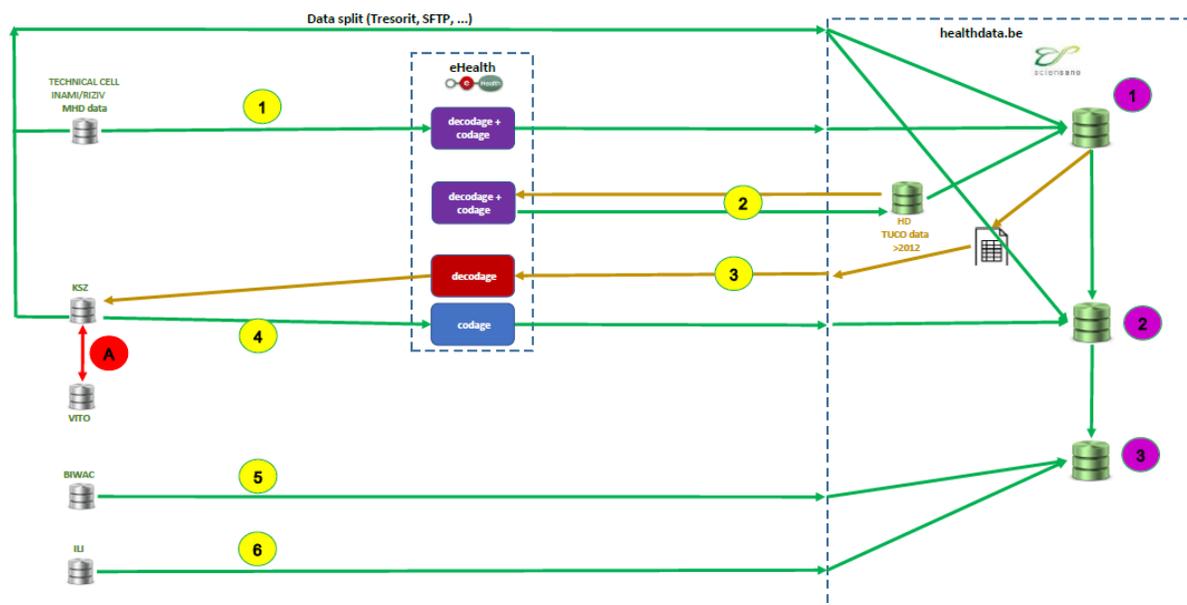
5) distance à l'axe routier majeur le plus proche : moyenne annuelle à la résolution GPS ;

6) *le code de la cellule de la cartographie RIO* comprenant les géocodes des adresses de résidence et aux adresses professionnelles. La cartographie RIO est une carte de la Belgique sur laquelle est appliquée une grille de cellules de 4x4km<sup>2</sup>. Chaque coordonnées GPS sera rattachée au code de sa cellule correspondante.

Le VITO fournira également des données non-environnementales provenant de cartographies disponibles via StatBel et Censtat et aux géocodes correspondant aux adresses de résidence et aux adresses professionnelles. Ces données non-environnementales seront de 2 catégories:

- 7) densité de population ;  
 8) indices socio-économiques de niveau d'éducation et de revenu.

8. Les données à caractère personnel pseudonymisées relatives à la santé seront collectées via 6 flux de données différents et trois points de traitement des données.



Explication des différents flux de données (cercles jaunes) :

- 1) Les données hospitalières minimales sont demandées à l'INAMI. Seuls les enregistrements pour une liste spécifiée de codes ICD seront demandés. Les données MHD seront transférées de la cellule technique de l'INAMI vers healthdata.be dans 2 fichiers distincts.
  - Un dossier contenant le NISS du patient avec un ID technique. Après l'encodage du NISS par le codage eHBox, les données sont transférées vers l'entrepôt de données healthdata.be.
  - L'autre fichier se compose des données MHD avec le même ID technique. Étant donné que ce deuxième fichier ne contient pas de données à coder, le fichier sera directement envoyé de la cellule technique à l'entrepôt de données de healthdata.be à l'aide de Tresorit, SFTP ou d'une autre méthode de transfert cryptée de fichiers.
- 2) Healthdata.be collecte les données TUCO depuis 2012. Pour les données TUCO, seul un identifiant de patient QERMID est actuellement présent dans l'entrepôt de données HD. Un recodage via l'équipe QERMID et ensuite via le décodage/codage eHBox est nécessaire pour obtenir un patient NISS codé 'healthdata.be' qui pourra être lié aux données MHD. L'ensemble de données lié des données TUCO et MHD sera utilisé comme données de base pour le projet EMIR.
- 3) Après avoir lié les données MHD et les données TUCO, une liste de patients est extraite pour laquelle des données supplémentaires sont demandées à la BCSS. Les NISS des

patients codés selon la clé 'healthdata.be, ainsi que la date de l'événement (infarctus) sont envoyés à l'eHBox pour décodage. La BCSS utilise NISS non codé pour récupérer les adresses et autres données des patients.

- 4) La BCSS se chargera de **coder les adresses en coordonnées GPS** en utilisant la base de donnée "Belgian Street and Adresses (BeSt) du Service Publique Fédérale Stratégie et Appui (FOD-BOSA). Le VITO se chargera de réaliser le lien avec les **données environnementales et non-environnementales** du système d'information géographique (GIS) en se basant uniquement sur le géocode, la date de l'infarctus et un l'ID technique Concernant les données environnementales avec variations journalières (température de l'air, PM 10, PM2.5; Ozone, NO2), le VITO limitera la fusion des données aux 45 jours précédents et aux 45 jours suivant l'infarctus. Cette procédure de liaison des données environnementales et non environnementales aux coordonnées GPS et à la date de l'infarctus se fera par le VITO.

**La liaison des données environnementales et non environnementales se fera dans les locaux de la BCSS et sous sa supervision.** Les données environnementales seront emmenées par le VITO à la BCSS pour effectuer la liaison avec les données de géolocalisation en utilisant les scripts python. Les données fusionnées ne seront stockées que sur le serveur de la BCSS. La BCSS se charge ensuite d'effacer les coordonnées GPS puis d'envoyer les données à healthdata.be via 2 fichiers distincts de la même manière que pour les données MHD (voir flux de données 1).

- 5) Les **données BIWAC** sont envoyées directement à healthdata.be. Ces données ne contiennent pas de données d'identification du patient et n'ont pas besoin d'être codées par la plateforme eHealth.
- 6) Les **données du registre ILI** sont collectées comme covariante supplémentaire dans l'étude. Les données ILI contiendront des informations géographiques sur la grippe en Belgique. Il s'agit de **données agrégées** aucun codage par eHealth n'est requis.

Points de traitement (cercles violets) :

- 1) Les 2 fichiers de données de MHD sont joints en fonction de l'ID technique présent dans les deux fichiers. Ensuite, les données MHD et les données TUCO sont jointes sur la base du NISS patient codé.
  - 2) Les 2 fichiers de données de BCSS / VITO sont joints en fonction de l'ID technique présent dans les deux fichiers. Ensuite, les données KSZ / VITO sont jointes à l'ensemble de données MHD / TUCO via le NISS patient codé.
  - 3) Les données BIWAC sont jointes avec les données MHD / TUCO / KSZ / VITO sur la base de plusieurs attributs (date d'admission, code d'accord hospitalier, sexe, âge à l'admission). Les données ILI seront jointes sur base du mandat géographique (code postal, district, ...).
9. Les données agrégées résultant de ce couplage seront analysées par les chercheurs de l'équipe EMIR regroupant des chercheurs de l'Université de Hasselt, de la VUB, de l'ULB. Les chercheurs n'auront pas accès aux données pseudonymisées utilisées pour réaliser le couplage.

## II. COMPÉTENCE

10. En vertu de l'article 42, § 2, 3<sup>o</sup>, de la loi du 13 décembre 2006 *portant dispositions diverses en matière de santé*, la chambre sécurité sociale et santé du Comité de sécurité de l'information est, en principe, compétente pour l'octroi d'une autorisation de principe concernant toute communication de données à caractère personnel relatives à la santé.
11. En vertu de l'article 5, § 1<sup>er</sup>, de la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, la Banque Carrefour de la sécurité sociale recueille des données à caractère personnel auprès des institutions de sécurité sociale, les enregistre, procède à leur agrégation et les communique aux personnes qui en ont besoin pour la réalisation de recherches pouvant être utiles à la connaissance, à la conception et à la gestion de la protection sociale.
12. En ce qui concerne la communication des données à caractère personnel du datawarehouse marché du travail et protection sociale, il s'agit par ailleurs d'une communication de données à caractère personnel qui, en vertu de l'article 15, § 1<sup>er</sup>, de la même loi du 15 janvier 1990 doit faire l'objet d'une délibération de la chambre sécurité sociale et santé du Comité de sécurité de l'information.
13. Le Comité s'estime dès lors compétent pour se prononcer sur la présente demande d'autorisation.

## III. EXAMEN

### A. ADMISSIBILITÉ

14. Le traitement de données à caractère personnel relatives à la santé est en principe interdit, et ce conformément au prescrit de l'article 9, §1<sup>er</sup> du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (RGPD).
15. Néanmoins, cette interdiction n'est pas d'application lorsque le traitement est nécessaire à des fins de recherche scientifique ou à des fins statistiques, conformément à l'article 89, paragraphe 1<sup>4</sup>, sur la base du droit de l'Union ou du droit d'un État membre qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée.

---

<sup>4</sup> « Le traitement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique, ou à des fins statistiques est soumis, conformément au présent règlement, à des garanties appropriées pour les droits et libertés de la personne concernée. Ces garanties garantissent la mise en place de mesures techniques et organisationnelles, en particulier pour assurer le respect du principe de minimisation des données. Ces mesures peuvent comprendre la pseudonymisation, dans la mesure où ces finalités peuvent être atteintes de cette manière. Chaque fois que ces finalités peuvent être atteintes par un traitement ultérieur ne permettant pas ou plus l'identification des personnes concernées, il convient de procéder de cette manière ».

16. En ce qui concerne le traitement de données à caractère personnel, le traitement n'est licite que si, et dans la mesure où, au moins une des conditions de l'article 6 du RGPD est remplie.
17. En vertu de l'article 5, § 1er, de la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, la Banque Carrefour de la sécurité sociale recueille des données à caractère personnel auprès des institutions de sécurité sociale, les enregistre, procède à leur agrégation et les communique aux personnes qui en ont besoin pour la réalisation de recherches pouvant être utiles à la connaissance, à la conception et à la gestion de la protection sociale.
18. En ce qui concerne la communication des données à caractère personnel du datawarehouse marché du travail et protection sociale, il s'agit par ailleurs d'une communication de données à caractère personnel qui, en vertu de l'article 15, § 1er, de la même loi du 15 janvier 1990 doit faire l'objet d'une délibération de la chambre sécurité sociale et santé du Comité de sécurité de l'information.
19. A la lumière de ce qui précède, le Comité est par conséquent d'avis qu'il existe un fondement admissible pour le traitement des données à caractère personnel pseudonymisées relatives à la santé envisagé.

## **B. PRINCIPES RELATIFS AU TRAITEMENT DES DONNÉES À CARACTÈRE PERSONNEL**

### **1. FINALITÉS**

20. Selon l'article 5 du RGPD, les données à caractère personnel doivent être traitées de manière licite, loyale et transparente au regard de la personne concernée. Elles doivent être collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités. Le traitement ultérieur à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques n'est pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales (limitation des finalités).
21. Il s'agit d'un traitement ultérieur à des fins de recherches scientifiques dans le cadre d'une étude épidémiologique sur la survenance de l'infarctus du myocarde en lien avec les conditions environnementales.
22. Au vu des objectifs du traitement tels que décrits ci-dessus, le Comité considère que le traitement des données à caractère personnel envisagé poursuit bien des finalités déterminées, explicites et légitimes.

### **2. MINIMISATION DES DONNÉES**

23. Selon l'article 5 du RGPD, les données à caractère personnel doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées.

24. Le Comité constate que les données à caractère personnel pseudonymisées suivantes seront traitées dans le cadre de cette étude. Les données sont considérées comme pseudonymisées car la récolte des données se fait soit sur la base du NISS (qui n'est pas communiqué au chercheur) ou sur la base de la date de l'infarctus. L'identité des patients concernés n'est pas communiquée au chercheur.

**a. Données issues de la base de données RHM – résumés hospitaliers minimum**

**1) MHD administratives**

***Date d'admission*** : Il s'agit d'une donnée maitresse pour l'étude EMIR et se justifie par la nécessité de projeter les données environnementales à un moment précis. Il sera ainsi effectué une analyse de la structure lag (relation entre les évènements d'infarctus et les caractéristiques environnementales des jours précédents) jusque 7 jours avant l'admission (lag -7).

***Année de naissance / Age à l'admission*** : D'un point de vue médical, l'âge est un facteur de risque cardiovasculaire qui peut modifier l'influence de l'environnement. Les personnes âgées sont plus sensibles aux particules fines, alors que les plus jeunes développent des infarctus en relation avec la pollution de l'air par le dioxyde d'azote. L'année de naissance servira également au croisement des données MHD avec les données anonymisées du registre BIWAC. En effet, l'âge et le sexe du patient associés à la date d'admission et au code d'agrément de l'hôpital d'admission permettront d'apparier les données afin d'obtenir une évaluation la plus précise possible du profil de risque cardiovasculaire des patients.

***Sexe*** : D'un point de vue médical, le sexe est un facteur de risque cardiovasculaire qui peut modifier l'influence de l'environnement. Les maladies cardiovasculaires sont plus fréquentes chez les hommes, mais le tako-tsubo affecte en grande majorité les femmes. Concernant les influences de l'environnement, l'influence du genre est encore indéterminée. Le sexe servira également au croisement des données MHD avec les données anonymisées du registre BIWAC.

***Code postal du patient*** : Le code postal est disponible dans les données MHD et sera avec le code de la cellule de la carte RIO, la seule donnée géographique encore présente dans la base de données finale accessible aux chercheurs. En effet les données d'adresse physique et professionnelle fournies par la BCSS à VITO seront effacées une fois les données environnementales complétées. Le code postal servira donc à établir des statistiques locorégionales de distribution des infarctus du myocarde mais, faute de résolution spatiale suffisante, ne pourront servir à modéliser les effets de la pollution de l'air.

***Numéro d'agrément de l'hôpital*** : le code de l'hôpital d'admission servira avec la date d'admission, l'âge du patient, son sexe et la date d'admission à permettre le croisement des données MHD avec les données anonymisées du registre BIWAC.

***Type d'admission*** : Le type d'admission des patients de l'étude EMIR doit consister essentiellement en des admissions urgentes. Ce paramètre servira à vérifier que les données ne sont pas contaminées par des erreurs d'encodage de patients admis électivement.

**Adressé par/destination** : beaucoup de patients STEMI sont pris en charge dans un premier centre et sont transférés dans un centre tertiaire pour réaliser une angioplastie avant de repartir dans leur centre référent. Ces paramètres serviront à identifier les patients ayant bénéficié de transferts secondaires. Dans le cadre des STEMI, le temps de transfert peut affecter le timing de la reperfusion et donc la sévérité de la présentation clinique. Il est donc important d'anticiper l'effet de ce cofacteur dans l'évaluation de l'effet propre à l'environnement sur la sévérité de la présentation initiale et la mortalité des infarctus.

**Type de sortie/ Date de sortie de l'hôpital/ Durée de séjour (date de sortie - date d'admission)** : le type de sortie vers domicile ou centre de réhabilitation ainsi que la durée totale de séjour sont aussi des indices indirects de la sévérité du tableau clinique qui doivent être appréhendés dans l'étude EMIR.

**Major diagnosis category/ Diagnosis related group/ Niveau de sévérité/ Risque de mortalité** : Les données MHD comportent, en plus de la nomenclature ICD du diagnostic principal, un code correspondant à une catégorie de diagnostic principal et un code de groupe de diagnostic. Ces codes sont associés à un calcul de niveau de sévérité de pathologie et de risque de mortalité attendu. Ces calculs de risque a priori se basent sur des données cliniques et comme l'étude EMIR comportera une évaluation réelle de la mortalité observée, il est important de déterminer si les influences de l'environnement peuvent expliquer les potentielles différences entre mortalité attendue et mortalité réelle.

## 2) MHD diagnostic principal

Il s'agit des données recoupant les différents **codes ICD** en diagnostic principal et permettant d'identifier les sous types de syndrome coronarien aigues entre STEMI, NSTEMI et Tako-Tsubo. Cette donnée majeure pour l'étude EMIR se justifie par la nécessité d'effectuer une analyse différentielle de l'influence de l'environnement sur ces 3 sous classes de syndromes coronariens aigus. En effet, leurs physiopathologies sont très différentes. Le STEMI est dû à une rupture de plaque de cholestérol, le NSTEMI d'un déséquilibre entre l'apport et la consommation d'oxygène par le myocarde, et le Tako-Tsubo à un problème de spasme coronaire microvasculaire probablement secondaire à un déséquilibre du système nerveux autonome. Les résultats préliminaires semblent suggérer que l'environnement affecte différemment ces sous types d'infarctus.

## 3) MHD diagnostic secondaire

Cette catégorie de données reprend une série de **code ICD** apparaissant dans les diagnostics secondaires et couvrant les facteurs de risque traditionnels de maladie cardiovasculaire. L'objectif est de couvrir le profil de risque cardiovasculaire du patient par les données présentes dans le RHM jointes à celles du QERMID-TUCO et à celles présentes dans le registre BIWAC. En effet, seul le croisement de ces différentes sources de données permettra une évaluation de qualité du risque cardiovasculaire. Ce croisement de données se justifie par le fait que les diagnostics secondaires du MHD n'ont pas pour mission principale de refléter les facteurs de risque cardiovasculaires mais ont l'avantage de couvrir toute la population de l'étude, tandis que les données QERMID et BIWAC ont quant à elles le but

d'évaluer précisément ce profil de risque, mais ne couvrent en revanche que les patients traités par stents (QERMID) ou ayant eu un STEMI (BIWAC). L'évaluation du profil de risque cardiovasculaire d'un patient se base sur l'identification de la présence d'un tabagisme, d'un diabète, d'une hypertension, d'une dyslipidémie et d'une insuffisance rénale chronique. Les patients d'ores et déjà connus pour avoir présenté une maladie vasculaire périphérique, une maladie cérébrovasculaire ou un problème coronarien chronique sont aussi d'office considérés comme à risque accru.

#### **4) MHD approche thérapeutique**

Les patients STEMI et NSTEMI bénéficient en très grande majorité d'un traitement par revascularisation coronaire selon 2 modalités : une revascularisation percutanée par mise en place d'un stent (PCI) ou une revascularisation par pontage (CABG). Il sera important de pouvoir donner une statistique sur ces modalités de revascularisation dans la description de la population incluse dans l'étude EMIR. En effet, la mortalité des infarctus en lien avec l'environnement ne peut s'interpréter que si tous les facteurs connus affectant cette mortalité sont appréhendés (profil de risque cardiovasculaire, sévérité initiale de la présentation et approche thérapeutique). De plus les données de PCI faisant l'objet de la base de données QERMID, il est important de vérifier la cohérence des données par des points de vérification. Ainsi tous les patients repris dans le QERMID devraient apparaître comme PCI dans le registre MHD. Cette cohérence entre les bases de données fera partie du processus d'évaluation de la qualité de la base de donnée finale.

#### **b. Données issues du registre Qermid TUCO**

##### **1) TUCO administrative**

Les justifications des données administratives du registre Qermid TUCO tel que le numéro d'agrément de l'hôpital, la date d'admission, l'âge à l'admission, le sexe du patient, l'année du décès du patient, le code postal sont les mêmes que pour les données du MHD.

La base de données Qermid TUCO reprend également la date de la PCI qui peut être différente de la date d'admission surtout pour les infarctus NSTEMI. En effet les recommandations pour les STEMI sont d'effectuer la PCI dans les 90 minutes suivant l'admission. Le temps entre l'admission et la PCI pour les infarctus NSTEMI est aussi un élément permettant d'évaluer indirectement la sévérité clinique de l'infarctus et l'étude EMIR devra utiliser ce paramètre dans la modélisation de l'influence de l'environnement sur la sévérité de la présentation initiale de l'infarctus.

##### **2) TUCO indication du traitement**

La base de données Qermid TUCO identifie clairement l'indication principale du stent. Cette donnée est essentielle à l'étude EMIR puisque seul les indications STEMI et NSTEMI seront retenues. En effet toutes les données concernant les PCI effectuées dans des situations stables ne seront pas retenues pour cette étude. Parfois l'opérateur renseigne en indication supplémentaire des informations qui peuvent être utiles à une meilleure description des cas inclus dans l'étude EMIR. Il s'agit principalement du type de PCI dans les STEMI et du délai de PCI pour les NSTEMI.

### **3) TUCO facteurs de risque cardiovasculaire**

La base de données Qermid TUCO, contient une évaluation assez précise du risque cardiovasculaire du patient (diabète, insuffisance rénale, maladie cérébrovasculaire (AVC), maladie vasculaire périphérique). De plus, le registre Qermid reprend des données de poids et de taille et donc l'indice de masse corporelle qui n'apparaissent pas dans les données MHD. En revanche, le tabagisme n'est pas repris dans les données Qermid TUCO. Cette situation illustre pourquoi les chercheurs ont opté pour une évaluation transversale du risque cardiovasculaire par le croisement des données MHD, Qermid TUCO et BIWAC. En effet seule la complémentarité de ces 3 bases de données permettra d'arriver à une évaluation fiable du risque cardiovasculaire des patients.

### **4) TUCO sévérité clinique**

L'étude EMIR se donne comme objectif d'établir si l'environnement influence la survenue des infarctus, mais aussi leur sévérité en terme de mortalité. L'évaluation de l'impact de l'environnement sur la sévérité initiale du tableau peut se faire que si un ensemble de facteurs cliniques est appréhendé. Le registre TUCO reprend clairement l'information de la présence d'un choc cardiogénique à l'admission, et si le patient a bénéficié d'un pontage chirurgical urgent, 2 facteurs signant un gravité de la présentation clinique. Enfin, le Qermid TUCO reprend les données de mortalité en précisant si la mortalité était bien cardiovasculaire, ou si un autre facteur a interféré (hémorragie, infection...) avec celle-ci. Ceci est important pour l'étude EMIR qui s'intéresse uniquement à la mortalité cardiovasculaire directement imputable à l'infarctus.

### **5) TUCO sévérité anatomique**

La mortalité doit également s'appréhender en fonction du degré d'atteinte coronaire et le registre Qermid TUCO permet d'avoir une évaluation fiable de cette dernière. En effet l'influence de l'environnement ne sera certainement pas la même chez un patient présentant une maladie n'affectant qu'un des 3 coronaire, plutôt qu'une maladie affectant d'emblée les 3 vaisseaux ou le tronc commun. Ainsi les données Qermid fournissent une description de la sévérité de la maladie (atteinte du tronc commun, atteinte de 1, 2 ou 3 vaisseaux, de la position proximale ou distale de la lésion traitée (numéro de segment), d'une lésion de bifurcation (nombre de stents), d'une atteinte du vaisseau dominant (dominance droite, gauche, co-dominance). De plus, le registre Qermid permet d'identifier les thromboses de stents qui sont une catégorie d'infarctus STEMI très particulière relevant plus d'une complication technique et qu'ils nous semblent donc important à identifier car pouvant fausser l'interprétation des influences de l'environnement sur les STEMI. De même, le registre Qermid permet d'identifier le nombre d'intervention durant l'hospitalisation. Ceci est important puisque seules les interventions primaires seront d'intérêt pour l'étude EMIR et non les éventuelles ré-interventions suite à des complications techniques. Enfin, ce registre comporte aussi des informations importantes à reporter dans la description des infarctus comme le type de stent utilisés.

c. **Données communiquées par la Banque Carrefour de la sécurité sociale (BCSS)**

*L'adresse de résidence principale du patient au moment de l'infarctus géocodée en données GPS* : cette donnée est capitale puisque l'étude EMIR se donne pour but d'analyser le biotope du patient, autrement dit l'environnement de son habitat en terme de température, qualité de l'air, exposition au bruit et degré de végétalisation de l'habitat. Il convient de préciser qu'il sera demandé l'adresse de résidence principale au moment de l'événement, qui peut donc différer de la dernière adresse connue du patient en cas de déménagement durant la période d'étude. D'autres études menées à l'échelle d'un hôpital ou d'une ville ont déjà travaillé sur les adresses de résidences des patients, mais l'étude EMIR sera la première à modéliser la pollution de l'air à l'adresse de vie grâce à un procédé d'interpolation. En effet, les études précédentes ont considéré la station de mesure de qualité de l'air la plus proche du domicile, mais souvent distante de plusieurs dizaines de kilomètres, comme étant une mesure fiable de la qualité de l'air de l'habitat.

Ces données de géolocalisations d'adresse ne seront plus présentes dans la base de données accessible aux chercheurs (niveau 2) et dans les publications scientifiques (niveau 3). Ces données GPS seront traitées par le VITO au sein et sous la supervision de la BCSS afin de les coupler aux paramètres environnementaux et non environnementaux correspondant. Une fois ce couplage effectuée, les données de géolocalisations seront effacées par la BCSS.

*L'adresse professionnelle au moment de l'infarctus géocodée en données GPS* : Aucune étude environnementale n'a pour l'instant pris en compte le temps passé en dehors de son domicile. L'étude EMIR se donne comme objectif secondaire de tempérer le niveau d'exposition entre le temps passé à domicile et le temps passé sur le lieu de travail. Il s'agira d'une modélisation mathématique pour pondérer de différentes manières le temps à domicile, le temps au travail et le temps de trajet domicile-travail puisque les temps réels d'exposition domicile/travail ne seront pas accessibles dans cette étude rétrospective. Comme pour les données de géolocalisations des adresses de résidences, ces données de niveau 1 seront supprimées par la BCSS et donc inaccessible au niveau 2 et 3.

*Données socioéconomiques* : l'influence environnementale ne peut s'appréhender que si les facteurs de risques cardiovasculaires établis sont évalués de manière fiable. A côté des facteurs médicaux tels que le tabagisme, le diabète etc, les facteurs socioéconomiques sont d'importants déterminants des maladies cardiovasculaires. Pour cette raison l'étude EMIR nécessite une évaluation de différents facteurs tels que le statut marital, la catégorie socio-professionnelle et la catégorie de revenu annuel. En effet, plusieurs études ont démontré que les personnes célibataires et/ou sans emploi et/ou au revenu modeste sont plus susceptibles de développer des maladies cardiovasculaires. Ces populations sont aussi plus à même de vivre dans des quartiers défavorisés, plus urbains et donc potentiellement plus pollués. Seule l'étude d'un large nombre de patients, et appréhendant à la fois des données sanitaires, socioéconomiques et environnementales permettra d'identifier les effets propres à l'environnement et quelles populations sont les plus à risque d'en développer des conséquences délétères.

*Statut vivant ou décédé, l'année du décès* : la survenue du décès du patient durant l'hospitalisation pour son infarctus est un des marqueurs principaux sur lequel va porter l'analyse de l'étude EMIR. En effet, au-delà du déclenchement de l'infarctus et d'une admission hospitalière, l'étude EMIR souhaite envisager si l'environnement aggrave la sévérité de la présentation de l'infarctus et donc sa mortalité. Certains patients peuvent décéder dans les semaines et mois qui suivent leur sortie de l'hôpital et donc l'étude EMIR portera sur la mortalité hospitalière, mais aussi la mortalité ultérieure (à 1 an post infarctus). Il est donc nécessaire d'accéder uniquement à l'année de décès pour les décès survenu en dehors du séjour hospitalier.

Le Comité constate que le chercheur ne demande que l'année de décès du patient concerné.

#### **d. Données issues du registre BIWAC**

Les données du registre BIWAC sont des données personnelles anonymisées et gérées par le groupe de travail de cardiologie aigüe de la société Belge de Cardiologie ne concernant que les patients ayant présenté un infarctus STEMI. Il couvre des données administratives, d'indication de traitement par PCI, de facteurs de risque cardiovasculaire, de sévérité clinique et anatomique similaire dans leurs justifications à celles de la base de données Qermid TUCO. Il s'agit de données très fiables mais limitées, car ne couvrant que les hôpitaux participant volontairement au registre. Nous espérons pouvoir améliorer la capture du profil de risque cardiovasculaire des patients et leur sévérité d'atteinte clinique et anatomique par la fusion des données du BIWAC avec celles du Qermid et du MHD. Comme cette fusion ne peut se faire à partir du code NISS absent de la base de données BIWAC, les chercheurs souhaitent croiser les données de ce registre avec celles des 2 autres uniquement sur la base de *la date d'admission, l'âge, le sexe du patient et le numéro d'agrément de l'hôpital d'admission*.

En plus des données Qermid, préalablement justifiées, le registre BIWAC évalue le tabagisme ; le score Kilip qui est une évaluation de la sévérité de la présentation clinique (dysfonction ventriculaire induite par l'infarctus) ; la survenue d'un arrêt cardiaque lors de la prise en charge qui impacte la sévérité clinique de l'infarctus ; l'heure de début des symptômes et l'heure de la reperfusion par la mise en place d'un stent. Ces heures permettent de calculer le temps total d'ischémie qui conditionne la sévérité de l'infarctus et sa mortalité.

#### **e. Données relatives aux infections respiratoires ILI**

Il s'agit ici de données agrégées reflétant à un niveau régional l'évolution des épidémies de gripes et autres infections respiratoires. Les 3 paramètres disponibles permettront d'évaluer l'intensité des épidémies de grippe saisonnière. En effet la grippe saisonnière est reprise comme un facteur déclencheur d'infarctus myocardique et doit être considérée comme un confondant potentiel des effets de la pollution particulaire qui sont eux aussi plus marqués en périodes hivernales.

#### **f. Données environnementales communiquées par VITO**

Il s'agit de données liées à des données personnelles (adresse de résidence et adresse professionnelle préalablement géocodées en coordonnées GPS). Les données demandées à VITO couvrent les facteurs environnementaux agissant à l'échelle de la population: pollution de l'air, température de l'air, exposition au bruit et végétalisation du milieu de vie. VITO accède aux coordonnées GPS codées par la BCSS et correspondant aux adresses de résidence et professionnelle au moment de l'infarctus. Pour chaque géocodes, le VITO fournira les données journalière moyenne des 45 jours précédents et des 45 jours suivant l'infarctus pour la température de l'air et les concentrations en polluant ainsi que les moyennes annuelles des autres paramètres des 8 années de l'étude. Les données environnementales survenant après l'évènement d'infarctus peuvent apparaître superflues mais sont nécessaires pour l'approche bidirectionnelles de l'analyse statistique en case crossover. Une autre approche statistique par machine learning explorera la dynamique complexe unissant la fluctuation temporelle des facteurs environnementaux et la fluctuation des taux régionaux et subrégionaux d'infarctus.

Pour ce type d'analyse les données d'infarctus seront donc agrégées à une résolution spatiale de 16 km<sup>2</sup> correspondant à la cartographie RIO basée sur une grille de cellules de 4\*4 km couvrant l'entièreté de la Belgique. Le VITO fournira donc le code de la cellule de la cartographie RIO incluant le géocode des adresses de résidence principales et professionnelles. Il sera impossible d'identifier un patient particulier à partir de ce code de cellule RIO puisque la cartographie de la Belgique est composée d'environ 1900 cellules de 16km<sup>2</sup> et que, sur base de la densité moyenne de population, 6000 habitants résident en moyenne dans chaque cellule. Les statisticiens pourront donc procéder aux analyses de données agrégées sur base de ce code de cellule RIO (données de niveau 2). En revanche, les données accessible au public (niveau 3) prendront uniquement la forme d'illustration de cartographie RIO avec un code couleur correspondant à l'incidence journalière d'infarctus par cellule.

25. Le Comité constate que les données seront accessibles en fonction de 3 niveaux :
- Niveau 1: données de base qui seront combinées et agrégées par les parties tierces de confiance (BCSS/eHealth) ;
  - Niveau 2: le dataset disponible pour les chercheurs (pseudonymisé, agrégé à un premier niveau) ;
  - Niveau 3: publication scientifique (complètement anonymisé).

### Accès aux données

i. Les **données de niveau 2** seront stockées et traitées par les outils statistiques présents sur le serveur d'Healthdata. Aucune donnée ne sera téléchargeable ou exportable sous n'importe quel format.

ii. Les **données de niveau 2** seront accessibles par 3 chercheurs ayant chacun leur propre code d'accès. Ces chercheurs signeront le non disclosure form.

iii. Concernant les **données temporelles**: Les données de niveau 2 comportent bien la date de l'évènement (infarctus) indispensable à une analyse épidémiologique. Toutes les autres données temporelles ont été limitées. Seule l'âge du patient et l'année de décès éventuel sera

accessible en niveau 2. Il sera renseigné en niveau 2 si le décès est survenu durant le séjour hospitalier.

iv. Concernant les **données de localisations**: Les données de niveau 2 ne comporteront plus de données de géolocalisation d'adresse de résidence ou d'adresse professionnelle.

Ce point est essentiel pour garantir l'absence de possibilité d'identification d'un patient ou d'un groupe de patient. Le point central est que les données de niveau 1 géolocalisant les patients seront traités par un tiers de confiance (VITO) au sein même de la BCSS. Le VITO se chargera de compléter les données environnementales et non environnementales aux coordonnées GPS. Pour l'analyse en case cross over, les données environnementales observées aux géocodes respectifs et couvrant les 45 jours précédant et les 45 jours suivant l'évènement seront établies. Une fois complétée, la BCSS se chargera d'effacer les données de géolocalisations d'adresse avant leur transfert au niveau 2. Les chercheurs de niveau 2 n'auront donc aucun accès aux données de géolocalisations d'adresse.

v. Concernant le **choix de la taille des cellules statistiques** pour l'analyse des données en niveau 2, nous précisons désormais plus clairement notre stratégie dans le formulaire de demande. Pour l'analyse en "machine learning", les données de géolocalisation seront agrégées à une résolution de 16 km<sup>2</sup>. Cette résolution est celle de la carte RIO qui est une carte de Belgique sur laquelle est appliquée une grille de cellules de 4x4 km.

Les chercheurs effectueront une analyse du taux journalier d'infarctus par cellule en fonction des variations concomitantes des caractéristiques environnementales de la cellule. Compte tenu de la taille de 16 km<sup>2</sup> de ces cellules statistiques RIO et de la densité moyenne de la population en Belgique, il ne sera pas possible pour les chercheurs de niveau 2 d'identifier un patient ou un groupe de patient particulier. En effet, plus de 1900 cellules RIO composent la carte de Belgique et une moyenne de 6000 personnes vivent dans chaque cellule.

vi. Concernant **les données de niveau 3** : Il s'agit essentiellement de statistiques descriptives, de statistiques inférentielles (odds ratio) et de random forest diagram établi par le "machine learning". A titre d'illustration dans les communications scientifiques et grand public, il sera publié des cartes RIO mais reprenant les incidences journalières d'infarctus en fonction des variations de qualité de l'air et de température. Ces cartes utiliseront une échelle de couleur par cellule de 16 km<sup>2</sup>.

26. Le Comité exige qu'une analyse de risque « small cells » soit effectuée sur les données réelles afin d'évaluer le risque de réidentification des patients concernés et notamment la taille des cellules géographiques envisagées pour l'étude.
27. Le Comité est d'avis, en vue de limiter les risques de réidentification de la personne concernée, que seule la date de survenance de l'infarctus peut être communiqué au chercheur. En ce qui concerne les dates exactes de séjour, le Comité suggère soit l'application d'une méthode statistique faisant appel à l'utilisation d'un delta, soit la communication de la durée de séjour.

28. A cet égard, le Comité rappelle que toutes les mesures nécessaires doivent être entreprises pour limiter le risque de réidentification des patients concernés. La publication des résultats ne peut donc pas comporter de cartes révélant le domicile de la personne concernée.
29. Le Comité estime que les données sont adéquates, pertinentes et strictement limitées au regard de l'étude envisagée.

### **3. LIMITATION DE LA CONSERVATION**

30. Selon l'article 5, §1er, e) du RGPD, les données à caractère personnel doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées. Les données à caractère personnel peuvent être conservées pour des durées plus longues dans la mesure où elles seront traitées exclusivement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques conformément à l'article 89, §1er, pour autant que soient mises en œuvre les mesures techniques et organisationnelles appropriées requises par le règlement afin de garantir les droits et libertés de la personne concernée (limitation de la conservation).
31. Les données à caractère personnel pseudonymisées seront conservées durant 10 ans à partir de la date de réception des données de l'année 2021 (qui ne sera possible qu'en 2023). Les données seront conservées sous une forme pseudonymisée au sein de la plateforme Healthdata.be. Les données pseudonymisées seront ensuite détruites.
32. Le Comité constate que cette durée de 10 ans est nécessaire pour répondre aux questions supplémentaires de la communauté scientifique suite à la publication des résultats de l'étude dans revues scientifiques internationales.
33. Le Comité estime que le délai de conservation envisagé est raisonnable au regard de l'étude envisagée. Par conséquent, le Comité autorise une conservation des données pseudonymisées jusqu'au 31 décembre 2033. Les données ne pourront ensuite être conservées que sous une forme strictement anonymisée.

### **4. TRANSPARENCE**

34. Lorsque les données à caractère personnel n'ont pas été collectées auprès de la personne concernée, le responsable du traitement est tenu de fournir les informations citées à l'article 14, §§1 et 2 du RGPD à la personne concernée.
35. Le traitement de données à caractère personnel envisagé est un traitement ultérieur de données à caractère personnel récoltées dans le cadre d'une finalité première autre que l'étude scientifique EMIR. Par conséquent, il n'y a pas de contact direct entre le processeur de l'analyse de données sélectionnées et pseudonymisées et les patients, et ce contrairement à une étude clinique sur des bases de données propre à un hôpital.

36. Le demandeur estime donc qu'il faudrait un effort disproportionné pour informer ces personnes. En effet, une information éclairée et l'obtention d'un consentement individuel représenteraient une charge de travail majeure compte tenu des 81297 patients attendus et de l'antériorité des événements (période d'étude s'étalant de 01/01/2012 au 31/12/2018). De plus certains patients seront décédés au décours de leur infarctus ou dans les suites d'autres affections. La mortalité intra-hospitalière des infarctus STEMI est ainsi d'environ 6% en Belgique et on peut donc estimer qu'entre 5000 et 8000 patients seront décédés au moment de l'exportation des données.
37. Enfin, du point de vue pratique, seule une communication écrite sur base des adresses des patients pourrait être envisagées complexifiant d'autant plus la sécurisation des adresses de résidence. De plus, le taux de réponse à ce type d'information écrite sont le plus souvent très faibles. L'étude EMIR repose sur une évaluation du risque absolu d'infarctus, ce qui sous-entend une couverture complète de la population et non l'analyse d'un sous-groupe de participants volontaires.
38. Dans un souci de transparence, les objectifs et la méthodologie de l'étude EMIR a d'ores et déjà fait l'objet d'une large communication auprès des pouvoirs publics, des communautés scientifiques cardiologiques nationales (Société belge de Cardiologie et Ligue Cardiologique Belge) et dans la presse médicale. La transparence vis-à-vis du grand public fera l'objet d'un site internet reprenant une description de l'étude et de ses résultats.
39. Le Comité est d'avis qu'il y a lieu de prendre toutes les mesures pour protéger les droits des personnes concernées y compris en rendant les informations publiquement disponibles sur un site web.
40. Le Comité est d'avis qu'il existe suffisamment de transparence quant au traitement de données envisagé.

## **5. SÉCURITÉ DE L'INFORMATION**

41. Conformément à l'article 5, §1er, f) du RGPD, les données à caractère personnel doivent être traitées de façon à garantir une sécurité appropriée des données à caractère personnel, y compris la protection contre le traitement non autorisé ou illicite et contre la perte, la destruction ou les dégâts d'origine accidentelle, à l'aide de mesures techniques ou organisationnelles appropriées (intégrité et confidentialité).
42. Pour garantir la confidentialité et la sécurité du traitement de données, tout organisme qui conserve, traite ou communique des données à caractère personnel est tenu de prendre des mesures dans les onze domaines d'action suivants liés à la sécurité de l'information: politique de sécurité; désignation d'un délégué à la protection des données; organisation et aspects humains de la sécurité (engagement de confidentialité du personnel, information et formations régulières du personnel sur le thème de la protection de la vie privée et sur les règles de sécurité); sécurité physique et de l'environnement; sécurisation des réseaux; sécurisation logique des accès et des réseaux; journalisation, traçage et analyse des accès;

surveillance, revue et maintenance; système de gestion des incidents de sécurité et de la continuité (systèmes de tolérance de panne, de sauvegarde, ...); documentation.

43. Le Comité constate que la BCSS, la plateforme eHealth, Healthdata et l'UZ Brussels ont désigné un délégué à la protection des données.
44. Le Comité constate qu'une analyse d'impact relative à la protection des données a été réalisée. Les résultats de cette analyse ont été communiqués.
45. Le Comité constate que l'échange de données mis en place fait appel à l'infrastructure de la plateforme Healthdata.be. Les échanges ont lieu conformément aux dispositions de la délibération n°15/009 du 17 février 2015, dernièrement modifiée le 3 mars 2020, relative à la méthode générique d'échange de données à caractère personnel codées et non codées relatives à la santé, dans le cadre de healthdata.be et de healthstat.be.
46. Le Comité constate que la BCSS sous-traitera la conversion des adresses privées et professionnelles des patients en coordonnées GPS à VITO, la BCSS ne disposant pas des outils nécessaires pour réaliser cette tâche.
47. Le Comité exige que le couplage des coordonnées GPS et des données environnementales ait lieu au sein des locaux de la BCSS.
48. Le Comité constate que toutes les chercheurs impliqués dans l'étude sont contractuellement tenus de respecter la confidentialité des données traitées.
49. Le Comité constate que l'équipe de chercheurs EMIR prévoit qu'un chercheur de l'Université d'Harvard sera impliqué en tant que consultant dans l'analyse de machine learning. A cet égard, le Comité rappelle qu'en application de l'article 45 du RGPD, aucune communication de données à caractère personnel pseudonymisées ne peut avoir lieu vers les USA en l'absence d'une décision d'adéquation. Le responsable de l'étude confirme que ce chercheur ne recevra aucune donnée, ni aucun autre chercheur d'une université belge ou étrangère. Aucun transfert de données ne sera par ailleurs possible hors du serveur d'Healthdata et toutes les analyses devront être effectuées en utilisant les logiciels mis à disposition sur leur serveur.
50. Même si cela n'est pas strictement requis par le RGPD, le Comité estime qu'il est préférable de traiter de telles données sous la responsabilité d'un médecin<sup>5</sup>. Le Comité rappelle que lors du traitement de données à caractère personnel, le professionnel des soins de santé ainsi que ses préposés ou mandataires sont soumis au secret.
51. Le Comité constate que le traitement de données sera effectué sous la surveillance et la responsabilité du médecin responsable de l'étude ainsi que du médecin responsable de Healthdata.

---

<sup>5</sup> Le Comité sectoriel a formulé cette préférence dans la délibération n° 07/034 du 4 septembre 2007 relative à la communication de données à caractère personnel au Centre fédéral d'expertise des soins de santé en vue de l'étude 2007-16-HSR « étude des mécanismes de financement possibles pour l'hôpital de jour gériatrique ».

- 52.** Le Comité rappelle qu'en vertu de l'article 35 du RGPD et de l'article 23 de la loi du 30 juillet 2018 *relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel*, une analyse d'impact relative à la protection des données doit, le cas échéant, être réalisée<sup>67</sup>. Le Comité constate que les résultats de cette analyse lui ont été communiqués.
- 53.** Le Comité rappelle qu'en vertu de l'article 9 de la loi du 30 juillet 2018 *relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel*, le responsable du traitement prend les mesures supplémentaires suivantes lors du traitement de données génétiques, biométriques ou des données concernant la santé :
- 1° les catégories de personnes ayant accès aux données à caractère personnel, sont désignées par le responsable du traitement ou, le cas échéant, par le sous-traitant, avec une description précise de leur fonction par rapport au traitement des données visées;
- 2° la liste des catégories des personnes ainsi désignées est tenue à la disposition de l'autorité de contrôle compétente par le responsable du traitement ou, le cas échéant, par le sous-traitant;
- 3° il veille à ce que les personnes désignées soient tenues, par une obligation légale ou statutaire, ou par une disposition contractuelle équivalente, au respect du caractère confidentiel des données visées.
- 54.** Le Comité attire l'attention sur les dispositions du Titre 6. Sanctions de la loi du 30 juillet 2018 *relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel*, qui prévoient des sanctions administratives et sanctions pénales sévères dans le chef du responsable du traitement et des sous-traitants en cas d'infraction aux conditions prévues dans le RGPD et dans la loi précitée du 30 juillet 2018.

---

<sup>6</sup><https://www.autoriteprotectiondonnees.be/publications/recommandation-n-01-2018.pdf>

<sup>7</sup><https://www.autoriteprotectiondonnees.be/publications/decision-n-01-2019-du-16-janvier-2019-disponible-en-neerlandais.pdf>

Par ces motifs,

**la chambre sécurité sociale et santé du comité de sécurité de l'information**

conclut que :

la communication des données à caractère personnel telle que décrite dans la présente délibération est autorisée moyennant le respect des mesures de protection de la vie privée qui ont été définies, en particulier les mesures en matière de limitation de la finalité, de minimisation des données, de limitation de la durée de conservation des données et de sécurité de l'information.

Michel DENEYER

Le siège de la chambre sécurité sociale et santé du comité de sécurité de l'information est établi dans les bureaux de la Banque Carrefour de la sécurité sociale, à l'adresse suivante: Quai de Willebroeck 38 - 1000 Bruxelles (tél. 32-2-741 83 11).