

Les statistiques sur les causes médicales de décès de A à Z

Table des matières

Α	LE CI	EPIDC, SES PARTENAIRES ET SES PUBLICS	3
A.1	L	E CEPIDC: IMPLANTATION, ROLE, VALEURS, SERVICES ET ORGANISATION	
	A.1.1	Le CépiDc	
	A.1.2	Le rôle du CépiDc	
	A.1.3	Valeurs et engagements	
	A.1.4	Les services	
	A.1.5	L'organisation	
A.2	L	ES PARTENAIRES DU CEPIDC	
	A.2.1	Les partenaires pour la production	{
	A.2.2	Les partenaires pour la diffusion	
	A.2.3	Les partenaires pour l'analyse	
A.3	L	ES PUBLICS DU CEPIDC	
	A.3.1	Les décideurs publics et agences sanitaires	
	A.3.2	Les équipes de recherche	
	A.3.3	Les médias et le grand public	
В	PRO	CESSUS DE PRODUCTION DES CAUSES MEDICALES DE DECES	15
_			
B.1		OLLECTE DES DONNEES	
	B.1.1	Le certificat de décès	
	B.1.1.1 Les différentes formes du certificat de décès		
		.1.1.1.1 L'origine du certificat : électronique ou papier	
	_	.1.1.1.2 La version du certificat 1997 ou 2017	
	В.1.1	.1.1.1.3 Le type de certificat : général ou néonatal	
		.1.1.2.1 Contenu du volet médical	
	_	.1.1.2.2 Volet médical complémentaire	
	_		



		Circuit de l'information	
	B.1.3	Saisie et numérisation des certificats papiers	20
		Suivi de la réception des volets médicaux	
	B.1.4.	Suivi de la transmission des certificats papiers	20
	B.1.4.		
	B.1.4.	3 Suivi de l'exhaustivité de la réception des volets médicaux par rapport aux décès Insee	21
3.2	TR	AITEMENT	22
	B.2.1	Contrôles et correction des données	22
	B.2.2	Appariement avec les données de l'Insee	24
	B.2.2.	1 Synchronisation Mensuelle	25
	B.2.2.	2 Synchronisation Annuelle	26
	B.2.3	Le codage des causes médicales de décès	27
	B.2.3.		
	B.2.3.		
		2.3.2.1 Etape 1 - Codage des expressions	
		2.3.2.2 Etape 2 - Identification de la cause initiale	
	B.2.3.		
		2.3.3.1 Le système de codage automatique : Iris	
		2.3.3.2 Le codagemanuel	
	B.2.3.	4 Les contrôles qualité du codage	
	B.2.4	Le circuit de codage à un certificat	33
C	LA DII	FFUSION, L'ACCES AUX DONNEES ET LEUR EXPLOITATION	34
C.1	Do	NNEES PUBLIQUES	24
J. ⊥			
	C.1.1	Les données ouvertes sur le site du CépiDc (Open Data)	
	C.1.2	Données intégrées au niveau national et international Les demandes de données non ré-identifiantes	
	C.1.3	ι ρς αρ <i></i> παηαρς αρ αρηρρος ηρη τρ-ιαρητιτιαητρς	2.0
2.2		CCES AUX DONNEES POUR DES FINALITES D'ETUDES ET DE RECHERCHE	36
J.2	C.2.1	CCES AUX DONNEES POUR DES FINALITES D'ETUDES ET DE RECHERCHE	36
2.2	C.2.1 C.2.2	L'enrichissement de cohortes et registres des maladies	36 36 36
2.2	C.2.1 C.2.2 C.2.3	L'enrichissement de cohortes et registres des maladies	36 36 36
	C.2.1 C.2.2 C.2.3 C.2.4	L'cers aux donnees pour des finalites d'etudes et de recherche	36 36 37 37
C.3	C.2.1 C.2.2 C.2.3 C.2.4	L'accès au SNDS La demande d'accès au SNDS EXPLOITATIONS ET TRAVAUX METHODOLOGIQUES DU CEPIDC	3636373838
	C.2.1 C.2.2 C.2.3 C.2.4	L'enrichissement de cohortes et registres des maladies L'enrichissement de cohortes et registres des maladies Le SNDS L'accès au SNDS La demande d'accès au SNDS SEXPLOITATIONS ET TRAVAUX METHODOLOGIQUES DU CEPIDC Les analyses du CépiDc	
	C.2.1 C.2.2 C.2.3 C.2.4	L'enrichissement de cohortes et registres des maladies Le SNDS La demande d'accès au SNDS EXPLOITATIONS ET TRAVAUX METHODOLOGIQUES DU CEPIDC Les analyses du CépiDc Les principales tendances de variation de la mortalité par cause depuis 2000	
	C.2.1 C.2.2 C.2.3 C.2.4 LE: C.3.1 C.3.1.	L'enrichissement de cohortes et registres des maladies Le SNDS L'accès au SNDS La demande d'accès au SNDS EXPLOITATIONS ET TRAVAUX METHODOLOGIQUES DU CEPIDC Les analyses du CépiDc 1 Les principales tendances de variation de la mortalité par cause depuis 2000 2 Les facteurs de variation de la mortalité par suicide	
	C.2.1 C.2.2 C.2.3 C.2.4 LE: C.3.1 C.3.1. C.3.1.	L'enrichissement de cohortes et registres des maladies Le SNDS L'accès au SNDS La demande d'accès au SNDS EXPLOITATIONS ET TRAVAUX METHODOLOGIQUES DU CEPIDC Les analyses du CépiDc 1 Les principales tendances de variation de la mortalité par cause depuis 2000 2 Les facteurs de variation de la mortalité par cause 3 Les inégalités sociales et spatiales de mortalité par cause	
	C.2.1 C.2.2 C.2.3 C.2.4 LE: C.3.1 C.3.1. C.3.1. C.3.1.	L'enrichissement de cohortes et registres des maladies Le SNDS L'accès au SNDS La demande d'accès au SNDS EXPLOITATIONS ET TRAVAUX METHODOLOGIQUES DU CEPIDC Les analyses du CépiDc 1 Les principales tendances de variation de la mortalité par cause depuis 2000 2 Les facteurs de variation de la mortalité par cause 3 Les inégalités sociales et spatiales de mortalité par cause	



A Le CépiDc, ses partenaires et ses publics

A.1 <u>Le CépiDc : implantation, rôle, valeurs, services et organisation</u>

A.1.1 Le CépiDc

Le Centre d'épidémiologie sur les causes médicales de décès (CépiDc) est une unité de service non mixte de l'Institut national de la santé et de la recherche médicale (Inserm)¹. Il remplit une mission légale de l'Inserm et réglementaire de la France de production de la statistique nationale des causes médicales de décès² :

- au niveau national : article L.2223-42 du Code Général des Collectivités Territoriales ;
- au niveau européen : règlement n° 328/2011, qui oblige les États membres à fournir les données à Eurostat dans les 24 mois suivant l'année de décès, en suivant des critères de cohérence (respect des référentiels de l'Organisation Mondiale de la Santé OMS) ;
- au niveau international : règlement des nomenclatures WHA20.18 adopté à la 20^{ème} Assemblée mondiale de la santé le 22 mai 1967, qui stipule que les États membres doivent produire les statistiques annuellement et appliquer toutes nouvelles révisions de la Classification Internationale des Maladies à la date prescrite par l'Assemblée.

L'intégration du CépiDc au sein d'un institut de recherche qu'est l'Inserm a permis au CépiDc d'être au plus près de la communauté de recherche en santé publique pour guider les développements et exploitations de la base de données de la façon la plus pertinente possible, instituant ainsi une interface réactive et experte. Ce positionnement permet d'être directement au contact des équipes les plus avancées en informatique médicale, en statistique appliquée à la médecine et en épidémiologie.

Depuis 2009, le CépiDc est rattaché à l'Institut de santé publique de l'Inserm qui, avec les 8 autres instituts thématiques de l'Inserm, assure, dans le champ d'intervention de l'Inserm, la coordination et l'organisation de la recherche par grands domaines, au-delà des activités des seules formations de recherche relevant directement de l'Institut.

L'Inserm est sous la double tutelle du ministère en charge de la santé et du ministère en charge de la recherche.

A.1.2 Le rôle du CépiDc

a. Le principal rôle du CépiDc est la production de la statistique nationale des causes médicales de décès (cf. A.1.1) qui consiste à :

² L'Inserm assure cette mission depuis 1968. Celle-ci était assurée par l'Insee avant cette date.



¹ Le CépiDc existe depuis 1998, suite à l'évolution du SC8 (service commun n°8) créé en 1979.

- recueillir les certificats de décès, les traiter et les coder conformément au standard de la classification internationale des maladies ;
- diffuser ces statistiques au niveau national (au grand public, aux chercheurs et à Santé publique France), européen (Eurostat) et international (OMS).

La fourniture des données françaises relatives aux causes médicales de décès à Eurostat par la France est requise par le règlement européen n°328/2011.

C'est au CépiDc que la France a confié cette mission. A ce titre, il bénéficie du statut d'autre autorité nationale (ONA pour *Other National Autority*) accordé par l'Institut national de statistique français, l'Insee. Il est ainsi positionné aux côtés du Service statistique public. En tant que partie prenante de la production de statistique publique, il rend compte à l'Autorité de la statistique publique.

Les données produites sont une clé de voûte de la statistique et de la recherche en santé publique.

Il s'agit d'une information de référence pour l'orientation des priorités de santé de publique et des thématiques de recherche associées. Elle est mobilisée pour le suivi des grandes interventions en santé publique et de l'état de santé de la population, pour les comparaisons temporelles et géographiques et les benchmarkings internationaux.

Par leur exhaustivité, ces données peuvent enrichir toutes les cohortes ou autres bases de données scientifiques ou administratives disposant de données identifiantes suffisantes, et ainsi permettre l'analyse de tous facteurs de variation de la mortalité.

La très grande majorité des problématiques de santé connues est évaluée à travers le nombre de décès observés, prédits ou estimés (vague de chaleur, alcool, tabac, grippe, Covid 19...). Par leur simplicité, les données de mortalité sont en effet aisément compréhensibles et le déploiement de la certification électronique (possibilité pour un médecin de certifier un décès via un site web) renforce la possibilité de donner une réponse rapide sur un sujet d'intérêt pour les médias et l'ensemble de la population.

La finalité ultime de la production des données des causes médicales de décès est leur appropriation et leur réutilisation par le grand public ou des équipes expertes. Ainsi, les données produites par le CépiDc, accompagnées de la documentation *ad hoc*, sont accessibles en ligne à travers un outil Open Data garantissant la confidentialité des données (cf.C.1.1), ou à toute équipe disposant de l'autorisation réglementaire associée, sur un environnement sécurisé (cf.C.2).

Le CépiDc contribue aux travaux internationaux en matière de production de données, participant :

- au Centre Collaborateur OMS (CCOMS) sur les classifications internationales en langue française, sur la partie relative à l'amélioration des règles et outils de production des données des



causes de décès.

- au Core group Iris (cf. encadré) : https://www.cepidc.inserm.fr/causes-medicales-de-deces/systeme-de-codage-automatique-iris.

Le système Iris

Iris est un système automatique qui permet de coder plusieurs causes de décès et de choisir la cause sous-jacente du décès.

Le CépiDc de l'Inserm a été l'un des pays fondateurs, avec l'Italie, la Hongrie et la Suède, de ce programme de codage automatique compatible avec la Classification internationale des maladies (CIM) et avec les systèmes de codage utilisés par la plupart des autres pays.

Iris facilite beaucoup la mise en œuvre de la Classification internationale des maladies (CIM) dont la complexité augmente sans cesse et permet d'améliorer la comparabilité internationale des données produites.

Il est basé sur le formulaire international de certificat de décès fourni par l'Organisation Mondiale de la Santé (OMS) et les causes de décès sont codées selon les règles de la CIM-10.

b. Etudes et recherches épidémiologiques sur les causes de décès

Le CépiDc contribue par son support à certaines études exploitant les données qu'il produit, en particulier si elles s'appuient sur une nouvelle méthodologie, recourent à des champs encore peu exploités ou si les données utilisées sont provisoires. L'intervention du CépiDc dans l'étude, sous forme de services ou de collaborations scientifiques, permet alors de limiter le risque d'interprétation fallacieuse de résultats dus à des artefacts de production.

Le CépiDc réalise de façon régulière des études sur des thématiques non spécifiques, notamment sur la distribution temporelle et spatiale de la mortalité par cause (cf.C.3.1). Enfin, en raison de l'urgence de certains sujets ou de l'absence d'équipes en capacité de répondre dans un calendrier contraint, il est parfois mandaté par la direction de l'Inserm ou les directions du ministère de la santé pour réaliser des études sur des thématiques plus spécifiques (alcool, qualité des soins, Covid-19, suicide...).

c. Travaux méthodologiques

En tant qu'unité de services dédiée à la connaissance en santé publique, le CépiDc doit autant que possible avoir recours aux méthodologies les plus rigoureuses dans le cadre de son activité de production, de diffusion ou d'exploitation des données.

Il lui faut parfois réaliser des adaptations de méthodologies existantes ou des développements



pour les appliquer à ses propres activités.

Parmi les sujets méthodologiques fréquents figurent la mise en œuvre de techniques de science des données pour la production réactive de données, la prise en compte des causes manquantes et des causes multiples, l'évaluation de la qualité et de la comparabilité des données, la conception de mesure d'inégalités socio-spatiales de mortalité ou les techniques de chaînage des données (cf. C.3.2).

A.1.3 Valeurs et engagements

Le CépiDc a pour objectif de fournir des données de qualité et un accompagnement à leur utilisation pour éclairer le débat public et la prise de décision en santé publique. On entend par donnée de qualité les données se conformant aux recommandations de l'OMS et d'Eurostat pour garantir parallèlement la précision, la réactivité et la comparabilité internationale. La finalité ultime de ses activités est de contribuer à l'amélioration de la santé des populations.

A.1.4 Les services

Outre la diffusion et la documentation des données qu'il produit, le CépiDc offre les services suivants :

- a. accompagnement dans la formulation d'un besoin spécifique de données, analyse du risque de réidentification,
- b. sous condition de l'obtention des autorisations réglementaires adéquates : extraction, appariement, mise à disposition des données dans un espace sécurisé (ex. : CASD).

Ces différents services sont mis en œuvre en collaboration avec l'Unité mixte de services France Cohortes. La répartition des rôles se résume ainsi :

Le CépiDc développe et transmet son expertise sur les données qu'il produit sur toutes les problématiques spécifiques à leur exploitation.

France Cohortes propose un cadre méthodologique et un support non spécifique aux données du CépiDc, elle anime la communauté des utilisateurs, dont le CépiDc, sur des thématiques méthodologiques communes (risque de réidentification, appariement, documentation, imputation).

En pratique, les équipes souhaitant accéder aux données du CépiDc passent par un portail commun à France Cohortes et un centre de services qui oriente les utilisateurs (accès via le Système national des données de santé, problématiques liées aux systèmes d'information et réglementaires) avant transmission de l'instruction de la demande spécifique au CépiDc.

France Cohortes est une unité mixte de services créée en 2021, proposant des services à des projets impliquant la production de données pour la recherche en santé publique, avec pour



principale cible les données des cohortes françaises. L'unité fournit des services adaptés à toutes les étapes du cycle de vie de la donnée :

- 1. mise à disposition d'un système d'information sécurisé permettant le recueil, la gestion et la mise en qualité des données, l'exploitation statistique, les extractions, la documentation et la valorisation des données,
 - 2. le support et la formation sur l'utilisation des outils proposés,
- 3. un accompagnement méthodologique sur la gestion des participants à une étude, les bonnes pratiques de data management et d'exploitation statistique,
- 4. un accompagnement réglementaire, notamment sur les formalités relatives aux appariements, à la sécurité du système d'information et la confidentialité des données,
 - 5. la gestion des accès des équipes de l'Inserm au SNDS.

A.1.5 L'organisation

Grégoire Rey est le Directeur du service, statisticien-épidémiologiste, docteur en épidémiologie et habilité à diriger des recherches.

Le CépiDc est organisé en deux pôles : le pôle de production des données et le pôle d'exploitation statistique.

Le pôle de production des données est sous la responsabilité de Diane Martin, responsable du pôle de production des données et de la démarche qualité au sein du service.

L'équipe est composée de :

- Zina Hebbache, responsable de l'équipe de codage (cf.B.2.3) et représentant la France au sein du Mortality Reference Group de l'OMS,
- les codeurs médicaux, qui assurent le codage de premier niveau non pris en charge automatiquement par le logiciel de codage automatique Iris,
- les nosologistes, qui assurent le codage expert des cas les plus complexes, et enrichissent le dictionnaire des termes rencontrés sur les certificats,
- Marie Frank, Médecin DIM des hôpitaux universitaires Paris Sud, qui apporte un support à la responsable de l'équipe de codage sur les cas les plus complexes,
- Aude Robert, informaticienne experte en traitement automatique du langage, assurant la veille technologique sur cette question, le développement d'outils d'automatisation du codage médical, mettant en œuvre et maintenant le processus d'appariement avec les données de l'état civil (cf. B.2.2),
- Rémy Flicoteaux, médecin DIM au siège de l'APHP, apportant un support à Aude Robert sur



les nouvelles technologies de traitement automatique des langues,

- une data manager mettant en place des indicateurs permettant de suivre le bon déroulement et l'évolution des flux de données.
- des techniciens « qualité », en charge du contrôle de la saisie des données non médicales (dates de naissance et de décès, communes de domicile et de décès,...) et du suivi de l'envoi/la réception des certificats.

Le pôle d'exploitation statistique est sous la responsabilité de Walid Ghosn, statisticienépidémiologiste. Il comprend un ingénieur informatique et statistique, en charge du développement d'outils d'Open Data et de documentation et de la prise en charge des demandes de diffusion et extraction des données. Ce pôle accueille les stagiaires, doctorants et post-doctorants en charge de réaliser des études ponctuelles.

Le CépiDc s'appuie en outre sur les services suivants de l'Inserm :

- le Département des systèmes d'information, en charge de la maîtrise d'œuvre des infrastructures et des applications, dont CertDc (cf. B.1.1.1.1),
- le Département des affaires juridiques pour le suivi des conventions de partenariats,
- la Délégation à la protection des données, pour le suivi de la conformité des traitements de données personnelles à la réglementation,
- l'Unité mixte de services France Cohortes (cf.A.1.4),
- la Délégation régionale Paris 11, qui prend en charge le suivi administratif, financier, et les ressources humaines du CépiDc.

A.2 Les partenaires du CépiDc

A.2.1 Les partenaires pour la production

Dans le cadre de la production de la statistique des causes de décès, plusieurs acteurs sont impliqués :

- Les médecins certificateurs, dont l'activité s'inscrit dans divers cadres institutionnels (ville, hôpital, maison de retraite...).
- La **Direction générale de la santé (DGS**), en charge du suivi et de l'évolution du cadre légal et opérationnel dans lequel s'inscrit la certification (notamment du format du certificat de décès et des modalités de transmission de ces informations à l'Inserm)3.
- Les communes, et plus particulièrement leur service d'état civil, chargé de réceptionner les certificats des décès survenus sur leur territoire au format papier, d'établir un avis d'état civil pour envoi à l'Insee et un bulletin d'état civil pour envoi à l'Agence régionale de santé (cf.

³ Une convention définit les responsabilités respectives de la DGS et l'Inserm dans la mise en œuvre de la certification électronique.



-

B.1.2.).

- Les Agences régionales de santé (ARS) qui, suivant le cadre légal actuel, doivent veiller à la remontée des certificats papier depuis les communes de décès, à l'ouverture puis à la transmission à l'Inserm.
- L'Insee avec qui le CépiDc réconcilie les données qu'il traite avec celles de de l'avis d'état civil4.
- L'Agence du numérique en santé (ANS) en charge du Centre de gestion des terminologies de santé, qui dirige depuis 2021 le Centre Collaborateur OMS, pour la maintenance de la Classification Internationale des Maladies en langue française. En son sein, le CépiDc collabore avec l'Agence du traitement de l'information hospitalière (ATIH).
- La Caisse nationale d'assurance maladie (Cnam), en charge de la centralisation des données du Système national des données de santé (SNDS), à qui le CépiDc transmet régulièrement les données des causes médicales de décès.

Enfin, la mise en place du SNDS instaure une gouvernance des données avec un pilotage stratégique par la **Direction de la recherche, des études et de l'évaluation** (Drees - Service Statistique du ministère en charge de la santé) et un comité des utilisateurs animé par le **Health Data Hub (HDH)**, susceptible de faire remonter des demandes d'évolution des bases sources, dont celles du CépiDc.

A.2.2 Les partenaires pour la diffusion

Dans le cadre de la diffusion de la statistique des causes de décès, les acteurs suivants sont impliqués :

- la Cnam, ainsi que le HDH, responsable de traitement du SNDS, en charge d'ouvrir les accès au SNDS central et de réaliser des appariements de bases de données avec le SNDS, dont les données des causes médicales de décès,
- Le **Centre d'accès sécurisé aux données** (CASD), acteur de mise à disposition des données dans un environnement conforme au référentiel de sécurité SNDS.

A.2.3 Les partenaires pour l'analyse

Le CépiDc réalise des travaux en coopération ou assure des expertises avec de nombreux partenaires de la recherche en santé publique ou en méthodes d'exploitation de données.

Ses partenaires sont des spécialistes du domaine de l'épidémiologie, de la géographie de la santé et des biostatistiques, mais également d'autres thématiques comme celle de l'intelligence

⁴ Ce partenariat est décrit par une convention.



Dernière mise à jour le 11/08/2022

artificielle. Il s'agit notamment :

- du groupe "Nutritional Methodology and Biostatistics" du Centre International de Recherche sur le Cancer (CIRC),
- du groupe Information Langue Écrite et Signée (ILES) du Laboratoire interdisciplinaire des sciences du numérique (LISN) du Centre national de la recherche scientifique (CNRS),
- de l'équipe "Méthodes statistiques de data-mining et apprentissage" du Centre d'études en informatique et communications du Conseil national des arts et métiers (Cnam),
- de l'équipe Health data- and model- driven Knowledge Acquisition (HeKA) Inserm-Inria.

A.3 Les publics du CépiDc

A.3.1 Les décideurs publics et agences sanitaires

En France, les acteurs en charge des politiques nationales de santé publique s'appuient aujourd'hui sur des agences sanitaires pour éclairer leur décision. En particulier, trois agences peuvent avoir intérêt à exploiter les données des causes de décès :

- Santé publique France : assure notamment la veille sanitaire et met en place des indicateurs d'alerte. En ce qui concerne l'alerte sanitaire, l'agence a principalement besoin de données très réactives, en quasi-temps réel, sans forcément souhaiter un recul historique ou une comparabilité internationale des données. C'est pourquoi elle s'intéresse particulièrement au déploiement de la certification électronique des décès, un flux de données permanent lui permettant d'exploiter ces données de façon automatisée.
- Elle vise également à produire des bilans post-crise à brève échéance, pour lesquels l'utilisation des données papier dès réception au CépiDc (cf.B.1.2) est également importante, sans qu'il soit nécessaire de disposer de données définitives, traitées conformément aux standards internationaux.
- Elle réalise des analyses de long terme, sur les données finalisées, qui permettent également de tester la cohérence des indicateurs très réactifs à partir d'un « gold standard ». Santé publique France et le CépiDc échangent très régulièrement pour assurer la meilleure adéquation entre les attentes et les services rendus.
- L'Agence nationale de sécurité des médicaments (ANSM), en charge de l'identification des effets indésirables des médicaments, qui relève de la pharmaco-épidémiologie. Elle peut utiliser les causes de décès pour identifier des potentiels effets indésirables spécifiques à la prise de médicaments, à travers les données du SNDS, conformes aux normes internationales. Elle souhaiterait un meilleur appariement des données au SNDS et une diminution des délais de production. L'ANSM aimerait également pouvoir identifier les effets



indésirables directement déclarés sur les certificats de décès, en particulier sur les certificats électroniques, mais n'a pas accès à cette information de façon réactive. Les échanges entre l'ANSM et le CépiDc ont lieu dans le cadre de projets d'analyse.

- La **Haute autorité de santé (HAS)**, en charge de l'évaluation de l'activité des établissements de soins et du suivi des effets à moyens et longs termes des effets des médicaments et dispositifs de soins. Son accès permanent au SNDS lui permet de traiter ces données.
- La Fédération nationale des observatoires régionaux de santé (Fnors) fournit des indicateurs de suivi de l'état de santé aux collectivités territoriales (en particulier les régions). Elle gère un outil d'Open Data appelé SCORE-Santé, qui publie notamment des indicateurs fondés sur les données des causes médicales de décès. Le CépiDc siège au comité de pilotage de SCORE-Santé.
- La **Drees**, en particulier à travers sa présidence de l'Observatoire national du suicide (ONS), exploite parfois les données en utilisant son accès permanent au SNDS. Le CépiDc est membre permanent de l'ONS.
- De façon plus ponctuelle, la **Cour des comptes** s'appuie parfois sur les données des causes de décès, via le SNDS, pour éclairer ses analyses.

Au niveau international les acteurs utilisant les données sont :

- **Eurostat**, qui produit régulièrement des benchmarks à partir des données européennes finalisées des causes médicales de décès,
- l'OMS, qui centralise les données nationales des causes médicales de décès de chaque pays en capacité de les produire dans le monde. Quelques études sont réalisées à partir de cette source, notamment par l'OCDE.

A.3.2 Les équipes de recherche

On peut distinguer principalement quatre types d'équipes de recherche et d'études selon leur usage des données :

Les équipes intéressées par une vision très populationnelle de l'évolution globale de la distribution des causes médicales de décès. Il s'agit le plus souvent d'équipes de démographes de l'Institut national des études démographiques (Ined) ou de géographes de la santé, notamment de l'Université de Paris Nanterre. Les équipes de démographie de l'Insee et de la Drees produisent également quelques études. Ces équipes sont très intéressées par un chaînage des données des causes de décès avec l'Echantillon Démographique Permanent (EDP), panel sous la responsabilité de l'Insee, qui existe avec un faible recul (depuis 2006) depuis la mise en place de l'EDP-Santé (chaînage de l'EDP avec le SNDS).



- Les équipes intéressées à mettre en évidence des associations causales entre des facteurs de risques environnementaux et contextuels ou individuels et la survenue de décès par cause spécifique. Ces équipes sont souvent issues des unités de recherche en santé publique de l'Inserm, mais peuvent également faire partie de l'Institut de radioprotection et de sûreté nucléaire (IRSN), l'Institut national de recherche et de sécurité pour la prévention des accidents du travail ou des maladies professionnelles (INRS), ou de l'Université Gustave Eiffel (ex-Ifsttar). Elles ont besoin de données individuelles fines pour les mettre en corrélation spatio-temporelle avec des facteurs contextuels, individuels à partir des données du SNDS ou en réalisant des appariements avec des études longitudinales (cohortes).
- Les équipes en charge de la production des registres de pathologies sur un territoire donné (typiquement un département). Ces équipes sont parfois affiliées à une unité de l'Inserm, et souvent à un Centre hospitalier universitaire (CHU). Dans un souci d'exhaustivité, ces équipes ont parfois besoin, lorsque leur pathologie d'intérêt est très léthale (comme l'AVC), de récupérer assez en amont les données des certificats de décès pour compléter leur registre. Le principal enjeu pour elles est d'obtenir un accès réactif aux données et de pouvoir les intégrer à leur système en respectant les contraintes de sécurité.
- Les équipes de cliniciens qui souhaitent avoir accès aux causes multiples de décès pour mieux analyser les processus morbides les plus fréquents relatifs à des pathologies spécifiques et assez rares. Ces équipes peuvent accéder aux données par le biais du SNDS.

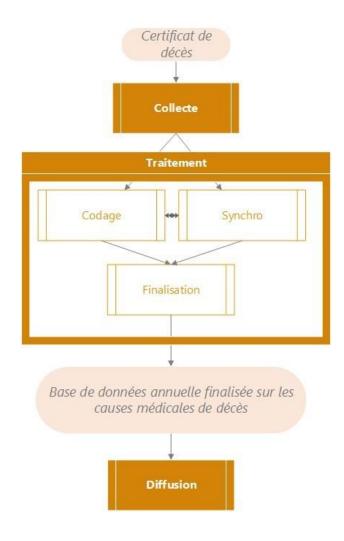
Il n'existe pas aujourd'hui d'instance permettant de recueillir et de centraliser des nouveaux besoins quant aux données sur les causes de décès.

A.3.3 Les médias et le grand public

Les médias, généralistes ou spécialisés, de même que le grand public, peuvent utiliser le site d'Open Data (cf. C.1.1) pour une vision synthétique de l'impact sanitaire d'une maladie, d'une circonstance de décès ou d'un évènement sanitaire sur la mortalité par cause. Le CépiDc répond dans la limite de ses compétences aux médias souhaitant accompagner les données d'un commentaire qualitatif, ou les oriente vers les experts du sujet. Il répond également aux demandes d'interviews en lien avec les reprises presse de certaines études grand public (analyse des grandes tendances).



B Processus de production des causes médicales de décès



B.1 Collecte des données

Lors de la survenue d'un décès sur le territoire français, un certificat de décès, document officiel indispensable aux démarches administratives pour la déclaration de décès et la gestion des corps, est établi. La production de la statistique nationale sur les causes médicales de décès s'appuie sur cette source de données administratives.

Le certificat de décès est composé de deux volets, tous deux rédigés par le médecin certifiant le décès de l'individu :

- le volet administratif comportant notamment le nom et les prénoms de la personne décédée, son adresse, la date et l'heure de son décès ;
- le volet médical, non nominatif, décrit au paragraphe « Le contenu du volet médical » (cf. B.1.1.2.1). C'est le volet médical qui constitue la source principale de données pour la production de la statistique nationale des causes de décès. Par abus de langage, le terme "certificat" peut être utilisé à la place de "volet médical" dans le cadre de la production des



données du CépiDc.

B.1.1 Le certificat de décès

B.1.1.1 Les différentes formes du certificat de décès

B.1.1.1.1 L'origine du certificat : électronique ou papier

Le certificat de décès peut être établi sur papier ou, depuis 2007, électroniquement via la plateforme CertDc, un site Internet sécurisé. Les informations sont identiques dans les deux formats.

La proportion de certificats électroniques augmente depuis la mise en place de la dématérialisation en 2007. Ainsi, entre mars 2020 et mi 2022, le taux de certification électronique est passé de 20 % à 34 %.

Le décret du 28 février 2022 renforce l'obligation d'établir les certificats de décès par voie électronique : elle s'applique aux décès ayant lieu dans un établissement de santé public ou privé ou dans un établissement médico-social, représentant environ 60 % de l'ensemble des décès. L'élaboration ou la transmission du certificat de décès sur support papier doit être, dans ce cas de figure, exceptionnelle : cette possibilité est limitée au cas où, pour des raisons techniques, le médecin certificateur se trouve dans l'impossibilité de transmettre un certificat de décès par voie électronique.

B.1.1.1.2 La version du certificat 1997 ou 2017

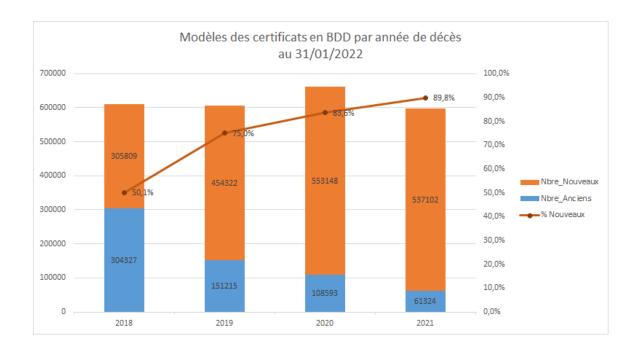
On distingue 2 versions de certificat de décès : celle de 1997 et celle de 2017, décrite dans l'arrêté du 17 Juillet 2017 et entrée en vigueur le 01/01/2018 pour les certificats au format papier ou électronique, qui devrait progressivement remplacer le certificat de 1997.

Pour les décès antérieurs à 2017, tous les certificats ont été établis dans la version de 1997. Pour l'année de décès 2017, 6 communes faisant partie de l'expérimentation de dématérialisation totale du certificat de décès ont utilisé la version 2017.

La dernière version contient davantage d'informations complémentaires, permettant de préciser les causes de décès. Elles ont été prises en compte dans le processus de traitement du CépiDc (cf. B.2) sans nécessiter la mise en place d'une double chaîne de codage des causes de décès (pour une année de décès donnée, les informations supplémentaires contenues dans la version 2017 du certificat sont simplement manquantes pour les certificats en version 1997).

L'évaluation de l'impact du nouveau modèle de certificats sur la statistique des causes médicales de décès est en cours au CépiDc.





B.1.1.1.3 Le type de certificat : général ou néonatal

Deux types de certificats de décès sont conformes à l'arrêté du 17 juillet 2017 en France :

- le certificat général ;
- le certificat néonatal.

Pour un décès concernant un individu d'au moins 28 jours, il convient d'utiliser le certificat général, tandis que les décès néonatals, concernant les enfants nés vivants (à plus de 22 semaines d'aménorrhée ou pesant plus de 500 g) et décédés au plus tard le 27ème jour de leur vie, doivent d'être retracés à l'aide d'un certificat néonatal.

En pratique, lorsque le médecin qui rédige le certificat de décès ne dispose que d'une version sur papier, il n'utilise pas toujours la version adéquate du certificat (général ou néonatal). C'est pourquoi le certificat de type général est parfois utilisé pour des décès néonatals et, plus rarement, le certificat de type néonatal peut être employé pour des décès d'adultes. Dans les deux cas, les volets médicaux correspondants doivent être traités manuellement, le logiciel de codage des causes de décès (cf. B.2.3) ne les prenant pas en charge de façon optimale. D'autre part, les décès de la période néonatale étant très sensibles, ils sont revus systématiquement. Les informations relatives à la mère de l'enfant, à l'accouchement ou au père, utiles à certaines études, sont par ailleurs manquantes dans le certificat général.

B.1.1.2 Le volet médical du certificat de décès

Le volet médical anonyme fait partie du certificat de décès utilisé comme document officiel indispensable aux démarches administratives pour la déclaration de décès et la gestion des corps. Il est rédigé par le médecin certifiant le décès d'un individu. Le volet médical constitue la source



principale de données pour la production de la statistique nationale des causes de décès. Par abus de langage, le terme "certificat" peut être utilisé à la place de "volet médical" dans le cadre de la production des données du CépiDc.

B.1.1.2.1 Contenu du volet médical

Le volet médical du certificat de type général comprend deux sections : celle contenant les causes de décès, dans laquelle le médecin certificateur donne l'explication médicale du décès qui paraît la plus probable et celle contenant les informations complémentaires en lien avec les causes de décès.

La partie sur les causes de décès, en texte libre, est elle-même divisée en deux parties.

La partie 1 contient l'enchaînement des évènements morbides ayant conduit à la mort en partant de la cause immédiate jusqu'à la cause initiale du décès.

La partie 2 contient des comorbidités, facteurs ou états physiologiques ayant contribué au décès mais non impliquées directement dans l'enchaînement des évènements décrits en partie 1.

Les informations complémentaires sont des cases à cocher sur les circonstances apparentes de décès (suicide, accident...), la mort subite, le lieu de décès, l'état de grossesse au moment du décès, l'activité professionnelle, et la demande éventuelle d'une recherche médicale ou médico-légale de la cause du décès.



INFORMATIONS RELATIVES AU DÉFUNT		nt constaté le décès – Renseignements confidentiels et anonymes
Commune de décès : Code p	ostal:	Date de décès :
Commune de domicile : Code p	ostal :	Date de naissance :
CAUSES DU DÉCÈS		
PARTIE I Maladie(s) ou affection(s) morbide(s) ayant dir Il s'agit de la maladie, du traumatisme, de l'intoxicatie (et non du mécanisme de décès comme une syncope, un a)		la complication ayant entraîné la mort processus morbide et le déc
due à ou consécutive à : b)		
due à ou consécutive à : C)		
due à ou consécutive à : d)		
La dernière ligne remplie dott correspondre à la cause initiale		es (grossesse) ayant contribué au décès, mais non mentionnés en Part
	GRO	OSSESSE La femme décédée était-elle enceinte ? non, pas au
☐ Domicile (du défunt ou autre) ☐ Établissement de santé p ☐ EHPAD, maison de retraite ☐ Établissement pénitentiaire	GRO	OSSESSE La femme décédée était-elle enceinte ? non, pas au pus au moment du décès, mais grossesse terminée depuis 42 jours ou moins oui, au moment du décès ne sait pas pas au moment du decès mais grossesse terminée depuis 91us de 42 jours et moins d'1 an
LIEU DU DÉCÈS Établissement de santé public Domicile (du défunt ou autre) Établissement de santé public EHPAD, maison de retraite Établissement pénitentiaire Voic publique Autre lieu ou indétermin MORT SUBITE S'agit-il d'un décès brutal et inattendu, évocateur de mort sul oui non ne sait pas * décès non troumatique (adulte, enfant, nourrisson) avec mode de survenue brutal (on moins d'une heure ou probablement) et inattendu (exclusion des maladies chroniques au stade terminal)	GR(□ ié □ bite*? La gr	DSSESSE La femme décèdée était-elle enceinte ? non, pas au cours de l'année précédant le décès depuis 42 jours ou moins ui, au moment du décès ne sait pas rossesses eat-t-elle contribué au décès ? non non ne sait pas non ne sait pas non ne sait pas
LIEU DU DÉCÈS Établissement de santé public Domicile (du défunt ou autre) Établissement de santé public EHPAD, maison de retraîte Établissement pénitentiaire Voic publique Autre lieu ou indétermin MORT SUBITE S'agit-il d'un décès brutal et inattendu, évocateur de mort sul oui non na le sait pas	GReat Grant	DSSESSE La femme décèdée était-elle enceinte ? non, pas au pas au moment du décès, pas au moment du décès, précèdant le décès depuis 42 jours ou moins depuis plus de 42 jours ou moins d'1 an moins d'une activité professionnelle décès est-il survenu lors d'une activité professionnelle des decès des des des des des des des des des de

La partie sur les causes de décès du certificat néonatal, bien que similaire, présente une spécificité : le médecin peut distinguer les causes maternelles des causes fœtales. Deux rubriques, l'une comportant les informations relatives à l'accouchement, l'autre celles relatives aux parents, enrichissent les informations complémentaires.



The state of the s		onstaté le décès – Renseignements confidentiels et anonymes (* instructions en annexe)		
INFORMATIONS RELATI	VES A L'ENFANT			
Commune de décès :	Code postal :	Date et heure de décès : à h Sexe : masculin		
Commune de domicile :	Code postal :	Date et heure de naissance* : i à h indétermin		
Apgar à 1 minute : Âge	gestationnel en semaines révolues d'an	nénorrhée : Poids de naissance en grammes :		
INFORMATIONS RELATIVE	S À L'ACCOUCHEMENT	INFORMATIONS RELATIVES AUX PARENTS (inscrire le code approp		
Naissance: 1. unique 2. gémellaire 3. trip Numéro d'ordre de l'enfant si grossesse multipl Lieu d'accouchement: 1. établissement de san Présentation: 1. sommet 2. autre céphalique Début du travail: 1. spontané 2. déclenché Mode d'accouchement*: 1. voie basse non ins 2. extraction instrumentale par voie basse 3.	té 2. domicile 3. autre 3. siège 4. autre 3. césarienne avant travail	MÈRE Année de naissance : Nationalité (en clair) : Profession* (en clair) : exercée pendant la grossesse : 1. oui 2. non 3. chômage 4. autre situation État matrimonial : 1. célibataire 2. mariée 3. veuve 4. divorcée La mère vit-elle en couple ? 1. oui 2. non Nombre total de grossesses, y compris grossesse pour cet enfant : Nombre total d'accouchements, y compris accouchement pour cet enfant* :		
Transfert ou hospitalisation particulière *de l'es	ıfant : 1. oui 2. non	PÈRE Profession* (en clair) : exercée pendant la grossesse : 1. oui 2. non 3. chômage 4. autre situation		
	e l'intoxication, de la complication ayant	(s) morbide(s) ayant directement provoqué le décès. t entraîné la mort (et non du mécanisme de décès comme une syncope, un arrêt cardiaqu		
CAUSE OBSTÉTRICALE OU MATE Autre(s) cause(s) obstétricale(s) ou mate	RNELLE* déterminante de la mort :			
INFORMATIONS COMPLÉMI	ENTAIRES (cocher la caze appropriée p	pour chaque point - *Lire les instructions de remplissage en annexe)		
LIEU DU DÉCÈS Établissement de santé public Établissement de santé privé	Domicile (du défunt ou autre) Voie publique Autre lieu ou indéterminé	RECHERCHE DE LA CAUSE DU DÉCÉS Une recherche de la cause du décès a-t-elle été demandée ? oui, recherche médicale oui, recherche médico-légale non son, un volet médical complémentaire sera établi ultérieurement par le médicen ayant réalisé le diagnostic des causes de d		
MORT INATTENDUE DU NOURRISSO oui	ne sait pas	SIGNATURE Nom lisible et cachet obligatoire du médecin		
The state of the s	Faits de guerre ns de soins médicaux, chirurgicaux			

Sont enfin renseignées sur les volets médicaux : la date de naissance, la date de décès, le sexe, la commune de décès et la commune de domicile. Ces informations, nommées par la suite « informations individuelles », sont utiles aux contrôles d'exhaustivité des décès et à l'enrichissement des données pour l'élaboration des statistiques sur les causes médicales de décès. (cf.B.2.2).

B.1.1.2.2 Volet médical complémentaire

Ce volet n'est destiné qu'aux personnes autorisées pour des motifs de santé publique (cf. art. L. 2223-42 du Code général des collectivités territoriales).

Lorsque les causes de décès ne sont pas connues au moment de la déclaration de décès, une recherche médicale ou médico-légale peut être demandée. Depuis 2018, conformément à l'article R. 2213-1-1 du Code Général des Collectivités Territoriales, le médecin ayant réalisé les investigations doit renseigner un volet médical complémentaire (VMC) via la plateforme de certification électronique, pour transmettre les causes de décès établies, quelle que soit la version de certificat de décès utilisé (1997 ou 2017). Le volet médical complémentaire est identique au volet médical (initial) du certificat de décès. Le VMC se substitue au volet médical (initial) s'il est plus complet. Sinon les informations sont fusionnées. Dans le cas où des informations sont contradictoires, c'est l'information du VMC qui est retenue. Cet arbitrage est réalisé dans le processus de traitement des certificats de décès.

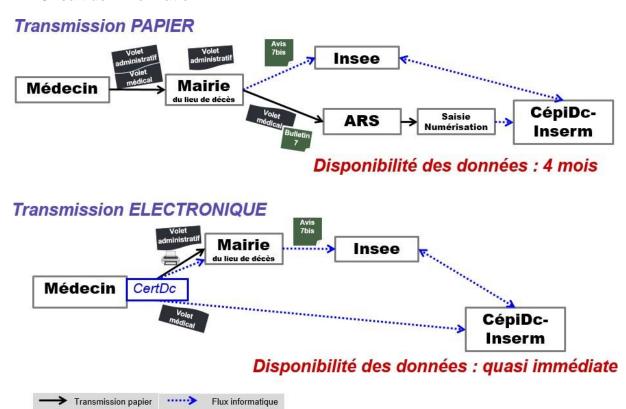


Bien que les VMC soient collectés via CertDc depuis 2018, peu ont été reçus, faute de communication : 100 pour l'année de décès 2018, 400 pour 2020 et 600 pour 2021. Il y aurait environ 10 000 autopsies par an en France.

Le CépiDc a mis en place une campagne dédiée courant 2020 auprès des instituts médicolégaux. Il en a résulté une légère augmentation du volume de VMC remontés mais toutes les morts suspectes donnant lieu à des recherches complémentaires ne font pas l'objet d'un VMC aujourd'hui.

Environ 2 000 VMC ont par ailleurs été récupérés directement via le système d'information de l'institut médico-légal de Paris. Les informations qu'ils contiennent, davantage issues d'une catégorisation simplifiée, ne comprennent pas d'enchaînement causal et s'avèrent moins riches, bien qu'utiles (attentats, suicides par exemple). Par ailleurs, les intégrer suppose à ce jour une exploitation manuelle. Les premiers constats (sur les VMC de 2018) font apparaître que le VMC n'est pas plus riche en informations que le volet médical initial. Seulement 33% d'entre eux ont été conservés quand le volet médical initial avait été reçu).

B.1.2 Circuit de l'information



Le médecin remplit les deux parties du certificat et scelle la partie médicale. En cas de certification papier, le certificat est transmis à la mairie. En cas de certification électronique, seule la partie administrative du certificat, imprimée ou dématérialisée, est transmise à la mairie. Si la mairie est raccordée au Hub d'Échange de l'État (HubEE), elle reçoit le volet administratif qui y est déposé par la plateforme CertDc de façon dématérialisée. Si elle ne l'est pas, le certificateur doit imprimer le volet administratif pour qu'il soit transmis à la mairie.

La mairie rédige alors deux documents :

- l'« avis 7 » bis, comportant le nom de la personne décédée et les informations d'état civil, envoyé à l'Institut national de la statistique et des études économiques (Insee), pour mise à jour du Répertoire national d'identification des personnes physiques (RNIPP);
- le « bulletin 7 », comprenant les mêmes informations sur la personne décédée, sans le nom, généré uniquement pour le circuit papier et transmis accompagné du volet médical toujours clos, à la délégation territoriale ou départementale de l'Agence Régionale de Santé (ARS), ou directement au siège, selon le fonctionnement régional. Le volet médical, ouvert par l'ARS, est ensuite envoyé à l'Institut national de la santé et de la recherche médicale (Inserm) accompagné du bulletin 7 via un prestataire de saisie et de numérisation.

En cas de certification électronique, le volet médical est transmis immédiatement au CépiDc par un transfert de données depuis la plateforme CertDc vers les serveurs et la base de données du CépiDc. Les traits d'identité (nom, prénom, date de décès, date de naissance, commune de décès) sont transmis automatiquement à l'Insee qui les utilise pour les présomptions de décès et dans le cadre de la synchronisation (cf B.2.2).

B.1.3 Saisie et numérisation des certificats papiers

Le CépiDc fait appel à un prestataire extérieur pour réceptionner les certificats de décès papier originaux, accompagnés des bulletins 7 en provenance des ARS. Ils sont numérisés et saisis selon des délais précis, puis les images scannées et les données saisies sont envoyées au CépiDc. Les causes de décès en texte libre font l'objet de nombreuses règles de saisie permettant de "nettoyer le texte" pour optimiser le codage des causes de décès (cf.B.2.3), en séparant les expressions médicales et en corrigeant les fautes d'orthographe. La séparation des expressions médicales permet de faciliter la reconnaissance du texte par les outils de traitement automatique des langues et ainsi optimiser le codage automatique de l'expression.

C'est pourquoi depuis les décès de l'année 2018, les causes de décès issues de la certification électronique sont également corrigées par le prestataire de saisie et numérisation des certificats papier (cf B.2.3.3.1).

B.1.4 Suivi de la réception des volets médicaux

B.1.4.1 Suivi de la transmission des certificats papiers

Chaque ARS (par délégation départementale ou territoriale ou de façon centralisée) doit envoyer avant le 10 de chaque mois tous les volets médicaux associés aux Bulletins 7 qu'elle a reçus le mois précédent de la part des mairies. Les mairies ont elles-mêmes 8 jours pour envoyer ces documents à l'ARS. Les instructions ministérielles des mairies et des ARS définissent la mise en œuvre de cette transmission. L'ARS doit notamment veiller à utiliser un mode d'envoi suivi et intégrer un bordereau de transmission à l'envoi. Ce bordereau doit être scanné et envoyé par mail au CépiDc afin de permettre à l'agent du CépiDc responsable du suivi de la transmission des certificats papier de vérifier l'arrivée du pli chez le prestataire de saisie et numérisation. Ce bordereau indique le



contenu du pli (nombre de certificats de décès par département concerné et mois concernés). Ces informations sont collectées et comparées aux indicateurs de réception fournis par ailleurs par le prestataire de saisie et numérisation. Ce processus permet d'identifier les envois perdus d'une part mais également d'identifier les ARS qui n'auraient pas fait leur envoi mensuel. Des relances sont alors faites chaque mois par l'agent du CépiDc aux ARS concernées.

B.1.4.2 Suivi de la transmission des certificats électroniques

Par définition, la transmission des certificats rédigés via l'application CertDc se fait de façon électronique via un dépôt de fichiers sur les serveurs du CépiDc. Un suivi du bon fonctionnement de transfert technique est effectué quasi quotidiennement.

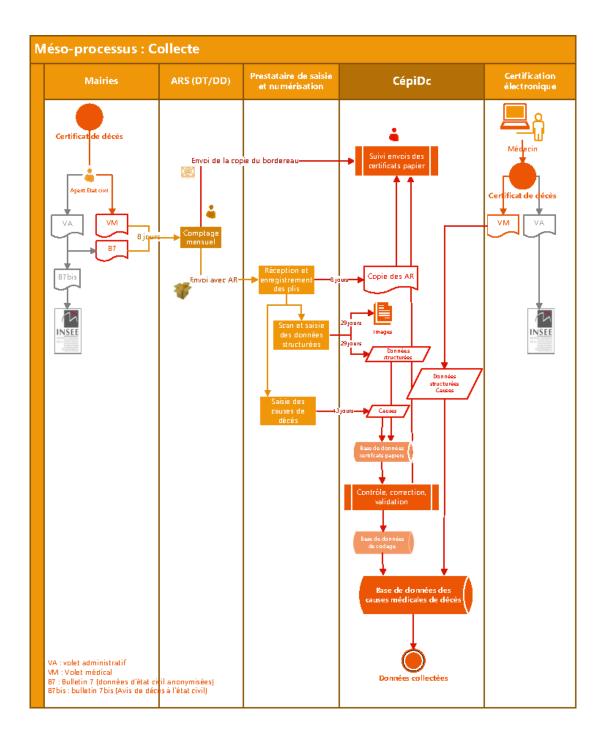
B.1.4.3 Suivi de l'exhaustivité de la réception des volets médicaux par rapport aux décès Insee

Le CépiDc a développé un tableau de bord lui permettant de suivre la réception des volets médicaux par rapport aux décès déclarés à l'Insee au fil de l'eau. L'Insee envoit les données brutes des avis 7bis reçus chaque mois au CépiDc. Ce sont ces données qui sont utilisées par le CépiDc pour construire ces indicateurs.

Tant que l'année de décès n'est pas terminée et que les chiffres de l'Insee ne sont pas consolidés, la proportion de volets médicaux reçus par rapport aux attendus n'est pas figée. L'Insee finalise les données d'une année de décès en juillet de l'année suivante.

Des certificats, notamment sous format papier peuvent parvenir au CépiDc plusieurs mois après le décès. Il arrive également que des reliquats parviennent l'année suivante. En moyenne, 97% des certificats de décès sont reçus dans les 4 mois suivant le mois de décès. Chaque année, pour 2% des décès, aucun volet médical n'est reçu au CépiDc.





B.2 Traitement

B.2.1 Contrôles et correction des données

Des contrôles de cohérence sont réalisés sur toutes les données structurées des certificats papier afin de vérifier la saisie du prestataire et corriger les données. Si une incohérence est détectée, des vérifications et corrections sont effectuées dans une application avec interface à partir de l'image numérisée. Les données structurées correspondent aux informations individuelles (sexe, date de décès, date de naissance, communes...) et aux informations complémentaires (cases cochées). Cette application permet de faire remonter les erreurs de saisie du prestataire et corriger les données.



Le CépiDc utilise notamment le bulletin 7 pour vérifier les informations individuelles présentes sur le volet médical papier. Si un écart est observé entre les données du bulletin 7 et les données du volet médical est observé, les données du bulletin 7 sont privilégiées. Si l'information est aberrante ou incohérente elle est corrigée avec l'autre document ou selon des règles de décision pré-définies. Si une donnée est manquante sur un des documents, celle indiquée sur le deuxième est utilisée. Les informations individuelles nécessaires au contrôle de l'exhaustivité des volets médicaux remontés (cf.B.2.2) sont particulièrement vérifiées : si elles sont manquantes, le certificat est également vérifié. Les mairies peuvent être recontactées si l'information est complètement manquante ou aberrante (année de décès dans le futur, date de décès antérieure à la date de naissance...).

Erreurs fréquentes observées :

Exemple 1 : La date de décès est identique à la date de naissance mais le certificat concerne une personne âgée (erreur fréquente du remplissage du certificat papier). Ces certificats sont à interpréter avec précaution : une mention de covid-19 sur un certificat de ce type ne concerne peut-être pas un nouveau-né.

Exemple 2 : Un décès du mois de janvier est daté de l'année précédente. On retrouve ainsi des volets médicaux avec mention de Covid-19 datés de janvier 2020 (avant le début de la crise sanitaire), enregistrés en janvier 2021. Une erreur de saisie du certificateur est alors fortement plausible. Le bulletin 7 peut permettre de corriger l'année de décès.

Des contrôles sont également mis en place pour vérifier la qualité de la saisie des causes (échantillons aléatoires). Il est à noter que le texte libre des causes de décès peut rester illisible pour certains certificats. La saisie à l'aide d'un logiciel de reconnaissance vocale peut également comporter des erreurs.

Les données du certificat électronique sont contrôlées à la saisie dans CertDc et une alerte est faite au besoin au médecin. Les contrôles sur la saisie des certificats électroniques doivent s'aligner sur les contrôles faits sur le papier : l'alignement est en cours dans le cadre de la refonte de CertDc.

Liste des contrôles donnant lieu à vérification des certificats papier :

- Données manquantes sur des données : Date de décès, date de naissance, commune de décès, commune de domicile, sexe
 - Date naissance > Date de décès
 - Date de décès > Date de réception du certificat décès
 - Année de décès < Année de réception du certificat -3
 - Année de décès = Année de Naissance et certificat de type adulte
 - Mois de décès = Janvier et année de décès < année de réception du certificat
 - Age au décès > 139 ans
- Ecarts entre donnée du bulletin 7 et du volet médical sur les données : Date de décès, date de naissance, code de la commune de décès, code du département du décès, sexe



- Si le modèle du certificat est néonatal :
 - Poids naissance <= 300 ou Poids naissance >= 5500
 - Nombre de grossesses > 20
 - Nombre d'accouchements > 20)
 - Age de la mère aberrant (inférieur à 9 ans ou supérieur à 60 ans)
 - Année de naissance de la mère manquante
 - Ordre de Naissance (rang) > 5
 - Score d'apgar à la naissance > 10
- Si modèle certificat est général, ces contrôles concernent des décès dits « sensibles » qui nécessitent une vérification exhaustive, sans que la cohérence des données soient remise en question : :
 - Une grossesse est mentionnée
 - Une contribution de grossesse au décès est mentionnée
 - Un accident du travail est évoqué

B.2.2 Appariement avec les données de l'Insee

Les informations individuelles contenues sur le volet médical éventuellement redressées après comparaison aux informations figurant sur le bulletin 7, pour la version papier du certificat (cf.B.1.1.1.1), sont ensuite exploitées pour apparier les données du volet médical aux données du Répertoire national d'identification des personnes physiques (RNIPP) géré par l'Insee (également appelées par la suite «données de l'Insee ». Aucun identifiant n'étant commun aux données du RNIPP et à celles du CépiDc à ce jour, on recourt à un appariement indirect.

Données Insee

Le répertoire national d'identification des personnes physiques (RNIPP) est tenu par l'Insee depuis 1946, et correspond à l'image des registres d'état civil. Il est mis à jour très régulièrement grâce aux bulletins statistiques de l'état civil établis et adressés à l'Insee par les communes à la suite de naissances, décès...Dans ce répertoire ne figurent que des informations sur l'état-civil : le nom, les prénoms, le sexe, la date et le lieu de naissance, la date et le lieu de décès pour les personnes décédées, le numéro de l'acte de naissance (et de décès) ainsi que le numéro d'inscription au répertoire (NIR). Le RNIPP sert pour la gestion de l'état civil et l'élaboration de statistiques démographiques. Chaque personne née en France y est identifiée une fois et une seule. Sa mise à jour régulière en fait la référence.

Le bulletin d'avis de décès que les communes envoient à l'Insee pour faire la déclaration de décès permet de mettre à jour le RNIPP. D'autre part, il contient des informations complémentaires non présentes dans le RNIPP permettant d'établir des statistiques démographiques (type de lieu de décès, activité professionnelle, le statut marital ...). Ces bulletins font l'objet d'un traitement spécifique et des redressements peuvent être opérés notamment sur les variables relatives à l'activité du défunt, sa profession, le lieu de domicile ou encore l'état matrimonial. Ce traitement permet d'assurer la



Cet appariement avec les données de l'Insee, appelé « synchronisation Insee » vise à :

- a. réaliser le suivi de la réception des volets médicaux reçus par le CépiDc et contrôler leur exhaustivité, afin de garantir la production de statistiques sur les causes médicales de décès portant sur le champ de l'ensemble des décès ayant eu lieu sur le territoire français déclarés à l'état civil;
- b. corriger éventuellement les données individuelles figurant sur le volet médical par les données consolidées du RNIPP, ce qui peut intervenir jusqu'au terme de l'opération de synchronisation (cf. Infra, pour assurer la cohérence statistique avec les autres sources de données sur le thème (statistiques sur les nombres de décès, données du Système national des données de santé). La répartition par âge, sexe, commune de décès seront donc identiques quelle que soit la source puisque ces données auront été mises en cohérence;
- c. enrichir la base de données du CépiDc par des données socio-démographiques fournies par l'Insee (état matrimonial, activité, profession, situation professionnelle, nationalité, pays ou commune de naissance (COG), commune de domicile (COG), lieu de décès déclaré à l'état civil).

B.2.2.1 Synchronisation Mensuelle

La synchronisation mensuelle peut concerner des certificats relatifs à différentes années de décès : un volet médical peut en effet arriver au CépiDc jusqu'à N+2 après le décès (N étant l'année de décès) et l'Insee intègre les actes de décès qu'il reçoit jusqu' à 6 mois après la fin de l'année N+1.

Chaque mois, les données des volets médicaux sont appariées par l'Insee avec les données des décès enregistrés à l'Insee. Si un volet médical est apparié avec un décès déclaré à l'état civil et enregistré au RNIPP, l'Insee renvoie au CépiDc les données individuelles enregistrées au RNIPP ainsi les données destinées à enrichir la base du CépiDc, listées ci-dessus. Les données du CépiDc sont donc enrichies et éventuellement corrigées : les date de naissance, date de décès, sexe, commune et département de décès, commune et département de domicile seront finalement identiques à celles de l'Insee au moment de l'appariement mensuel.

L'appariement mensuel couvre plusieurs années de décès (jusqu'à N-2).

Les données d'appariement utilisées par l'Insee pour la synchronisation mensuelle diffèrent selon l'origine du certificat :

 Pour les certificats papiers : le CépiDc envoie à l'Insee les données indirectement identifiantes au fur et à mesure de la réception : date de décès, date de naissance, commune de décès, commune de domicile, sexe, numéro d'acte de décès.



 Pour les certificats électroniques : l'Insee reçoit directement depuis la plateforme CertDc les données permettant un appariement plus fiable avec le nom et le prénom en sus des dates de décès, date de naissance, commune de décès, commune de domicile, sexe.

Les volets médicaux non appariés par l'Insee ne sont pas renvoyés au CépiDc (qu'ils aient été reçus depuis l'application électronique CertDc ou depuis le CépiDc pour les données issues de format papier) De la même manière, les décès Insee non appariés ne sont pas envoyés au CépiDc dans le renvoi mensuel. Ils restent dans la base qui sera servira à une nouvelle tentative d'appariement le mois suivant.

B.2.2.2 Synchronisation Annuelle

Contrairement à la synchronisation mensuelle, la synchronisation annuelle porte sur une unique année de décès.

Pour une année de décès donnée, une fois la réception des données considérée achevée à l'Insee comme au CépiDc , une « synchronisation annuelle » est réalisée au CépiDc. A ce jour, la synchronisation annuelle est réalisée à N+2 de l'année de décès.

La phase préliminaire à la synchronisation annuelle est la réception des fichiers Insee : l'Insee fournit l'ensemble des données relatives à l'année de décès, déjà fournies lors des synchronisations mensuelles (ne sont envoyées que les données relatives à des volets médicaux présents lors de la synchronisation mensuelle). Ces fichiers contiennent les identifiants Insee et Inserm pour les décès appariés en synchronisation mensuelle ainsi que les identifiants des décès Insee non appariés. Les identifiants sont bien sûr associés à l'ensemble des données dont le CépiDc a besoin (données d'appariement et d'enrichissement de la base décrites plus haut). En cas d'anomalie connue, l'Insee peut éventuellement refaire tourner l'algorithme d'appariement mensuel sur l'ensemble des décès de l'année concernée avant d'envoyer les fichiers annuels finalisés au CépiDc.

Le CépiDc s'assure ensuite que les appariements sont corrects et tente de réaliser un ultime appariement, pour les décès non ou mal appariés de part et d'autre. Il peut arriver que la synchronisation annuelle vienne modifier la synchronisation mensuelle.

L'objectif de la synchronisation annuelle est d'optimiser le nombre de décès Insee avec des causes médicales reçues au CépiDc. En effet, dans le cas où, pour un décès Insee, le CépiDc ne retrouve pas le volet médical dans ses bases, la cause de décès produite sera indéterminée. Pour ces cas-là, soit le volet médical n'a jamais été reçu, soit les données d'appariement étaient trop différentes. De la même manière, des volets médicaux reçus au CépiDc mais dont les données d'appariement ne correspondent à aucun décès Insee, seront supprimés de la base du CépiDc. Il s'agira soit de doublons, soit les données d'appariement étaient trop éloignées ce qui correspond à une perte de données car le décès Insee aura une cause indéterminée.

Actuellement, 98% des volets médicaux envoyés à l'Insee sont appariés dans la synchronisation mensuelle. 2 à 3 % de ces appariements semblent incohérents (écarts significatifs entre données). Lors de la synchronisation annuelle ils sont désappariés pour être retentés sur



l'ensemble des décès de l'année. Il peut arriver qu'un volet médical soit apparié avec un mauvais décès Insee en mensuel quand le "vrai" décès Insee est reçu côté Insee tardivement. Cela est rattrapé dans la synchronisation annuelle. La synchronisation annuelle permet d'affiner la synchronisation mensuelle avec des arbitrages sur le contenu des certificats. Une application avec interface permet de comparer les données des volets médicaux et des décès Insee afin de choisir le meilleur appariement, notamment quand il existe des doublons ou que l'appariement est "souple" (écart entre les données d'appariement).

Le « dédoublonnage » définitif est effectué lors de ces arbitrages : il s'agit, lorsque le CépiDc a reçu plusieurs certificats pour un même décès, de retenir celui apportant le plus d'informations sur les causes de décès. Deux volets médicaux présentant une date de décès, une date de naissance, un sexe et une commune de décès identiques sont considérés comme doublons. Si l'un d'eux correspond à un volet médical complémentaire et que le contenu sur les causes médicales de décès est plus informatif, les données qu'il contient sont prises en compte. Sinon, sont retenues les données du volet médical le plus pertinent. A pertinence égale, c'est le dernier volet médical qui sera conservé lors de la synchronisation annuelle. Le dédoublonnage s'effectue lors de la synchronisation annuelle puisqu'il permet également d'arbitrer quel volet médical sera apparié au décès Insee (et ce n'est pas forcément celui qui l'a été en synchronisation mensuelle).

A l'issue de cette seconde phase de la synchronisation, dans la base de données du CépiDc, un et un seul volet médical est associé à chaque décès déclaré à l'Insee et les données individuelles sont définitives. Pour les volets médicaux non reçus au CépiDc, un volet médical 'vierge' est généré : les causes de décès seront alors indéterminées. Les volets médicaux non appariés à un décès Insee sont supprimés de la base de données du CépiDc.

Enfin, des volets médicaux correspondant à des enfants sans vie peuvent être reçus au CépiDc. Ils sont identifiés à l'aide des causes de décès. Ces situations sont en théorie déclarées à l'Insee par des bulletins d'état civil spécifiques et ne doivent pas faire l'objet d'un bulletin de décès. Lors de la synchronisation annuelle, ils sont donc supprimés des bases de données du CépiDc. Cependant, pour certains, le processus administratif de déclaration de naissance et donc de décès a été mis en place et l'Insee a bien enregistré ces situations comme des décès. Ils sont donc conservés dans le processus statistique des décès côté Insee et le CépiDc conserve alors le volet médical dans ses données. Cette dernière situation peut donner lieu à des doublons dans le comptage des enfants morts-nés entre les données produites par la Drees et celles produites par le CépiDc.

B.2.3 Le codage des causes médicales de décès

La phase suivante du processus de production de la statistique sur les causes de décès correspond à la codification des causes de décès mentionnées sur le volet médical pour identifier la cause initiale définie comme la maladie ou le traumatisme qui a déclenché l'évolution morbide conduisant directement au décès ou, en cas de mort violente, les circonstances de l'accident ou de la violence qui ont entraîné le traumatisme mortel.



B.2.3.1 La nomenclature utilisée

Le codage s'appuie sur la Classification internationale des maladies (CIM), la classification médicale proposée et recommandée par l'Organisation mondiale de la santé (OMS) pour le codage en morbi-mortalité, et mondialement utilisée pour l'enregistrement des causes médicales de morbidité et de mortalité.

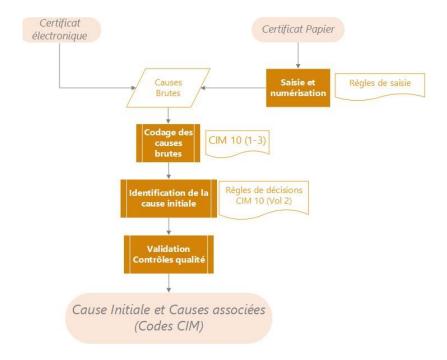
Elle permet de classer les maladies mais également les signes, symptômes, lésions traumatiques, empoisonnements, circonstances sociales et causes externes de blessures ou de maladies. Elle indique aussi comment sélectionner la cause initiale du décès selon l'enchaînement et le contexte des causes de décès renseignées sur le volet médical.

La version utilisée actuellement est la CIM 10. Elle sera bientôt remplacée par la CIM 11. La CIM 10 comporte environ 12 000 codes et a fait l'objet d'une publication sous la forme de trois volumes :

- Volume 1 : table analytique contenant la classification à 3 et à 4 caractères, la classification de la morphologie des tumeurs, des listes spéciales pour la mise en tableaux pour la mortalité et la morbidité, les définitions et le règlement.
- Volume 2 : manuel d'instruction du codage, en particulier en mortalité pour identifier la cause initiale des décès. Il contient l'ensemble des règles de décisions à appliquer pour coder les causes de décès et sélectionner la cause initiale.
- Volume 3 : index alphabétique. Cet index est l'entrée indispensable de la classification. Les termes courants et moins courants des maladies y sont répertoriés et associés au code CIM 10.

C'est la classification internationale qui définit le format des données nécessaires à la production des données sur les causes médicales de décès. Le format du volet médical français est conforme : il recueille les informations nécessaires pour appliquer les recommandations de l'OMS.





Le codage des certificats est effectué en deux étapes :

- Codage des expressions : l'attribution d'un code à chaque maladie, traumatisme ou cause externe de décès mentionné sur chaque ligne du certificat
- Identification de la cause initiale : par application des règles de décisions définies par l'OMS.

B.2.3.2 Les deux étapes du codage

B.2.3.2.1 Etape 1 - Codage des expressions

Cette première étape, essentielle, consiste à attribuer un code CIM à chaque terme, expression, cause médicale présente sur le volet médical en partie 1 et 2. C'est à partir du codage des expressions que l'identification de la cause initiale se fera en deuxième étape. C'est pourquoi la qualité de cette étape est primordiale.

Il ne s'agit pas de prendre les expressions indépendamment les unes des autres et de les rechercher dans le dictionnaire. En effet, le code à choisir pour une expression peut dépendre d'une autre expression présente ailleurs sur le certificat, selon leur lien causal. Il peut également dépendre d'une information fournie sur le contexte du décès comme le caractère accidentel ou suicidaire, l'état de grossesse chez une femme, ou encore l'âge du défunt. Les codes utilisés en périnatalité ne sont pas toujours les mêmes que les codes utilisés pour les adultes, y compris pour une même expression. Ces informations de contexte peuvent être précisées par le certificateur dans le texte ou de façon plus structurée dans les informations complémentaires sous forme de cases cochées. Quand une même expression peut être codée différemment selon le contexte, elle est appelée "expression à choix de code".

Pour réaliser la première étape de codage, on utilise le dictionnaire francophone CIM 10 qui



correspond au volume 3 de la CIM pour la francophonie. Ce dictionnaire a été créé par le CépiDc, qui en assure la maintenance. Il s'agit d'une table composée d'une ligne par libellé de cause associée à son code CIM.

Certaines expressions sont associées à 2 codes complémentaires. Les expressions dites à "choix de code" décrites ci-dessus sont identifiables dans le dictionnaire : elles font l'objet de plusieurs lignes pour la même expression (avec les différents codes).

B.2.3.2.2 Etape 2 - Identification de la cause initiale

Les règles de sélection et de modification définies dans le volume 2 de la CIM permettent d'identifier la cause initiale du décès à partir des codes des causes multiples, en prenant en compte le contexte et leur lien causal. Ces règles sont rassemblées sous forme d'organigramme dans le manuel d'instruction de la CIM pour être appliquées simplement et de façon homogène par les codeurs.

Ces règles sont appliquées dans un ordre précis, et permettent de guider la réflexion pour aboutir à la cause initiale.

a. Les règles de sélection

Ces règles permettent au codeur de sélectionner une cause initiale en respectant le plus possible les informations rapportées par le médecin certificateur en partie 1 dans un premier temps. La cause initiale mentionnée par le médecin sur le certificat peut être ambiguë, erronée ou ne pas répondre aux besoins statistiques. Par exemple, le médecin peut : mentionner plus d'une cause initiale (ambiguïté) ; remplir le certificat à l'envers ou donner un enchaînement causal improbable (par exemple un infarctus à l'origine d'une grippe) ; mentionner une cause initiale parfaitement acceptable, comme une dépression entraînant un suicide, mais la présentation des statistiques de mortalité selon la seule cause initiale privilégie la sélection de la cause externe (suicide) par rapport à la maladie (dépression).

Les règles de sélection prennent en compte l'enchaînement causal décrit par le certificateur selon qu'il soit probable ou non.

b. Les règles de modification

Une fois que les règles de sélection ont été appliquées, les règles de modification interviennent pour s'assurer que la cause initiale sélectionnée est la plus pertinente en prenant en compte l'entiereté du certificat. Une cause initiale identifiée par les règles de sélection peut être remplacée dans le cas où les règles de modification interviennent : il peut s'agir par exemple d'une cause initiale non acceptable (règle de la "cause improbable" selon la CIM), les règles de modification indiquent alors quelle cause initiale il faut choisir (à l'aide de tables de décisions par code). Il peut également y avoir, ailleurs dans le certificat (notamment en partie 2) une pathologie qui paraît de façon évidente comme à l'origine du processus morbide décrit en partie 1 (règle de la "cause évidente") et la CIM impose alors de la sélectionner. Ainsi, Les règles de modification de la cause



initiale permettent de choisir la cause initiale de décès, même lorsque le certificat est mal renseigné, ou permettent de préciser le diagnostic.

B.2.3.3 Les systèmes de codage

B.2.3.3.1 Le système de codage automatique : Iris

On observe des différences de codage importantes entre codeurs, au niveau national ou entre pays. Pour assurer la qualité et la comparabilité internationale des données produites, mais aussi pour des questions d'optimisation de ressources, de plus en plus de pays optent pour les systèmes de codage automatiques, comme le système Iris, compatible avec la CIM et utilisé par la plupart des pays. Le CépiDc en a été un des fondateurs.

Le système Iris facilite beaucoup la mise en œuvre de la Classification internationale des maladies (CIM) dont la complexité augmente régulièrement. Les codeurs experts traitent manuellement les cas complexes, non codés automatiquement, à l'aide de l'interface du système Iris. Le système Iris attribue des codes aux causes multiples à l'aide du dictionnaire, puis identifie la cause initiale du décès par application des règles de décision.

Bien qu'en constante évolution, le système Iris atteint ses limites lorsque le lien entre les causes multiples est complexe ou lorsque le contexte du certificat rend le codage des expressions difficile.

Pour certains cas, l'OMS demande explicitement un regard humain. Le logiciel remonte alors le certificat dans l'interface pour qu'une validation manuelle soit effectuée.

Pour le codage de l'année de décès 2017, 58 % des certificats de décès papier étaient codés automatiquement, contre seulement 41 % des certificats électroniques. L'écart était dû au traitement qu'effectue le prestataire de saisie des certificats papiers (cf.B.1.3). Sur les certificats électroniques, les expressions étaient moins « nettoyées » : seules quelques correctifs mineurs liés aux accents sont appliqués.

L'outil de codage automatique rejette de ce fait dès la première étape de codage davantage de certificats électroniques que de certificats pour le papier. A partir des décès 2018, la saisie des certificats électroniques a été reprise en appliquant les mêmes règles de saisie que pour les certificats papier, ce qui a permis d'augmenter la performance du codage automatique sur les certificats électroniques (58% également).

D'autre part, une maintenance du dictionnaire du CépiDc, notamment sur les expressions à choix de code a permis d'augmenter le codage automatique à 64% pour l'ensemble des certificats des décès de 2018 et ultérieurs.

Les rejets, liés notamment à des problèmes de syntaxe, des fautes d'orthographe, à la complexité de l'expression ou encore à l'absence de séparateurs, se produisant à l'étape 1 du codage automatique, constituent la grande majorité des rejets. Pour les certificats papiers, 13 % des rejets



contiennent des mots qui n'ont pas été déchiffrés par le prestataire de saisie. En version électronique, il n'y a pas de problème de lecture de l'écriture de médecin. Une autre partie des rejets est liée aux expressions à choix de code qui nécessite une interprétation humaine avec l'entièreté du certificat.

Certains certificats sont enfin rejetés en étape 2 car le lien causal entre les causes multiples est trop complexe ou trop ambigu ou parce que le système de codage automatique détecte une incohérence dans le certificat.

Méthode de traitement des mots illisibles dans les certificats

Pour les années de décès après 2017, un programme de codage automatique a été mis en place pour traiter les certificats contenant des mots illisibles dans les certificats papier. Deux tentatives de lecture sont effectuées : au moment de la saisie par le prestataire et lors du codage manuel. Si la lecture ne permet pas de lever l'imprécision après analyse par un codeur, l'expression est ignorée pour le codage (toute la ligne ou uniquement l'expression). Ce programme permet de coder automatiquement 83 % des certificats illisibles. Les certificats restants sont vérifiés par les codeurs experts et les termes illisibles sont également ignorés pour le codage manuel.

B.2.3.3.2 Le codage manuel

L'interface du système Iris permet de réaliser un codage manuel en retouchant le texte des certificats, en attribuant les codes CIM et en appliquant les règles pour identifier la cause initiale. Le codage peut être entièrement manuel mais le codage automatique peut être activé à tout moment via l'interface. Ainsi, pour les certificats bloqués à l'étape 1 pour lesquels l'expression médicale utilisée par le médecin contenait une simple erreur de syntaxe, le codeur peut "nettoyer" le texte et demander le codage automatique.

On distingue trois niveaux de codage manuel :

- lorsque le problème se situe à l'étape 1 et qu'il s'agit de problème de syntaxe ou d'orthographe, on parle de "pré-codage";
- lorsque le problème se situe à l'étape 1 et qu'il s'agit d'un problème de formulation ou d'une légère complexité pour appliquer les règles à l'étape 2, on parle de codage. Les codeurs expérimentés sont appelés "nosologistes". Ils maîtrisent les règles de codage et connaissent les expressions médicales courantes.
- lorsque le problème concerne des cas cliniques rares et complexes et que des connaissances en physiopathologie sont nécessaires, on parle d'expertise médicale. Les experts ont une expérience médicale et maîtrisent les règles de codage complexes.

B.2.3.4 Les contrôles qualité du codage

Pour s'assurer de la qualité du codage, des contrôles sur certaines catégories de certificats codés sont réalisés de façon exhaustives. Les certificats sont tous vérifiés quel que soit le mode de



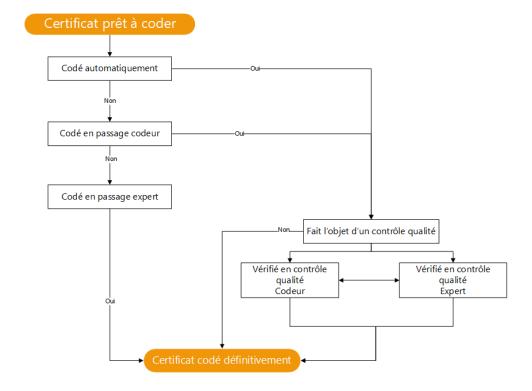
codage (automatique ou manuel).

Une vérification manuelle est alors faite par l'équipe de codage via l'interface du logiciel Iris pour valider ou corriger le traitement du certificat.

Ces contrôles ont différents objectifs :

- Répondre aux exigences Eurostat ou OMS
 - o Identifier les incohérences de codage
 - o Repérer les causes de décès non acceptables selon les règles de codage
 - o Déceler les incohérences entre l'âge ou le sexe et les causes
 - Vérifier certaines causes (morts violentes, morts maternelles, néonatales, covid-19...)
- Palier aux problèmes techniques identifiés dans le programme de codage automatique (le codage automatique pourrait être incorrect)
- Assurer une qualité du codage
 - o Bien coder les décès identifiés comme sensibles (mortalité infantile, SIDA...)
 - S'assurer de l'homogénéité du codage manuel, notamment pour certaines règles de codage complexes ou nouvelles.

B.2.4 Le circuit de codage d'un certificat





C La diffusion, l'accès aux données et leur exploitation

C.1 <u>Données publiques</u>

C.1.1 Les données ouvertes sur le site du CépiDc (Open Data)

Le vecteur de diffusion de données privilégié par le CépiDc est l'Open Data : un maximum d'informations sont diffusées sous cette forme.

Plusieurs indicateurs sont proposés sur le site principal Open Data (https://opendata-cepidc.inserm.fr/) mis en ligne en 2022 :

- les effectifs de décès, qui indiquent les grands volumes de décès,
- les **taux bruts de mortalité**, qui donnent le nombre de décès par unité de population domiciliée (source Insee), la taille de la population d'une unité d'analyse étant le premier facteur à contrôler pour interpréter la distribution spatio-temporelle des décès,
- les taux standardisés de mortalité, qui estiment le taux de mortalité sous l'hypothèse que la structure de la population par âge et sexe suit une distribution standard. Cet indicateur permet de contrôler les différences de distribution spatio-temporelle de la population par âge et sexe, et ainsi d'interpréter les variations des indicateurs après prise en compte de ses principaux facteurs de variation connus.

Le site permet de requêter ces indicateurs par département/région, âge et sexe et cause de décès. La ventilation par cause de décès est proposée selon les 86 catégories de la liste regroupée élaborée par Eurostat en 2012 pour limiter la sensibilité des tendances aux évolutions de codage tout en conservant le maximum d'information signifiante. Toutefois, une prochaine évolution de ce site prévoit la possibilité de ventiler les données par cause plus précise, après analyse du risque de réidentification et mise en place de mesures préventives (exemples : regroupement de catégories, floutage des données).

Dans cette attente, l'ancien site Open Data est toujours en ligne (http://cepidc-data.inserm.fr/inserm/html/index2.htm) et propose de faire des requêtes identiques, par cause plus fine, avec un regroupement plus ancien des causes de décès et des populations standards moins à jour.

Open Data Covid-19

Suite aux nombreuses demandes des médias sur le nombre de décès liés à la Covid-19 estimé à partir des certificats de décès, le CépiDc a développé un site dédié à la crise sanitaire en 2020 : https://opendata.idf.inserm.fr/cepidc/covid-19/

Ce site indique la répartition des décès avec mention de Covid-19 identifiés par le biais d'un



algorithme automatique par département/région, âge et sexe, lieu de décès, jour/semaine/mois.

Pour tenir compte de l'avancement de la réception et du traitement des données, et permettre ainsi à l'utilisateur d'apprécier leur couverture et leur exactitude, le site permet de distinguer :

- Les données issues des seuls certificats électroniques, reçues en quasi-temps réel, mais non représentatives de la totalité des décès, en raison d'un déploiement inégal de la certification électronique (cf. B.1.1.1.1.) selon le lieu de décès et le département de décès. Ces données permettent de suivre de façon assez fine les variations de la mortalité, avec quatre jours de délais.
- Les données cumulant les données des certificats de décès électroniques et papier, dont plus de 95 % sont reçues dans les 4 mois suivant le décès. Ces données sont provisoires, car issues d'un traitement automatisé, mais sont déjà d'une fiabilité suffisante pour donner une vision précise de la distribution spatiale et temporelle des décès.
- Les données codées de façon définitive par les experts du CépiDc. Elles permettent d'obtenir des statistiques suivant les recommandations de l'OMS, en ne retenant que les décès pour lesquels la cause initiale retenue est la Covid-19.

Enfin, le site fournit des indicateurs de suivi permettant de se situer dans la dynamique de remontée des données.

Les données fournies par ces deux sites peuvent répondre aux questions des médias, du grand public, et des utilisateurs plus experts souhaitant porter un premier regard sur les données.

C.1.2 Données intégrées au niveau national et international

Les données sur les causes de décès alimentent les productions d'autres organismes, complétant l'information produite par le CépiDc.

Au niveau national:

- Toutes les modifications de la base des causes médicales de décès sont transmises en quasitemps réel à Santé publique France pour lui permettre de remplir ses missions de veille et d'alerte sanitaire (cf. A.3.1). Elles lui permettent de produire des bilans hebdomadaires proposant une analyse de l'évolution de la mortalité en période de grippe ou de crise sanitaire mais également de réunir l'ensemble des indicateurs de veille sanitaire sur ses différents sites Internet, dont : https://geodes.santepubliquefrance.fr/
- Le site de la Fnors SCORE-Santé (https://www.scoresante.org/) propose également des indicateurs (effectifs, taux standardisés) de mortalité par cause à une échelle départementale, au fil de la transmission annuelle par le CépiDc. Ces indicateurs peuvent être mis au regard d'autres indicateurs de suivi de l'état de santé des populations en France, issues notamment du SNDS.



Au niveau international:

- Eurostat propose sur son site (https://ec.europa.eu/eurostat/fr/web/health/data/database) des indicateurs (effectifs, taux standardisé) de mortalité par cause à l'échelle régionale. Ces indicateurs sont disponibles pour tous les pays européens et permettent ainsi de faire des comparaisons par regroupement de causes de décès en 86 catégories. Il faut noter qu'Eurostat peut réconcilier les décès survenus en Europe, hors du pays de domicile des personnes, apportant ainsi une plus-value aux données produites au niveau national.
- L'OMS propose également sur son site (<u>https://www.euro.who.int/fr/home</u>) des indicateurs de mortalité par cause, pour tous les pays en capacité de les produire.

C.1.3 Les demandes de données non ré-identifiantes

Dans le cas où les différents sites produisant des indicateurs de mortalité, dont ceux proposés par le CépiDc, ne peuvent répondre à une question (trop de requêtes successives nécessaires, granularité spatio-temporelle spécifique, variables non proposées en Open Data), il est possible à toute personne de déposer une demande de données sur le site du CépiDc (onglet « Accéder aux données du CépiDc ou du SNDS »).

NB : Après analyse du risque de réidentification, le CépiDc peut orienter le demandeur sur une demande de données personnelles soumises à des exigences de sécurité supplémentaires.

C.2 L'accès aux données pour des finalités d'études et de recherche

L'accès aux données les plus fines des causes médicales de décès doit être uniquement voué à des fins d'études ou de recherche, à condition de respecter toutes les contraintes réglementaires et de sécurité informatique nécessaires. Ces données peuvent être exhaustives sur la population générale, ciblées sur une population en particulier ou encore enrichir des cohortes et registres des maladies. Les principales finalités d'exploitation des données sont exposées en C.2.3

Il est possible d'accéder aux données personnelles, soit par le biais d'une extraction et d'une mise à disposition dans un environnement sécurisé conforme au référentiel de sécurité SNDS, soit directement en accédant au SNDS par le portail sécurisé géré par la Cnam.

C.2.1 L'enrichissement de cohortes et registres des maladies

Après vérification par le centre de services de la conformité des autorisations transmises délivrées par la Cnil, l'extraction des causes médicales de décès est réalisée par appariement avec une liste de variables identifiantes communiquées via le tiers de confiance (Décret no 98-37 du 16 janvier 1998). Le résultat est finalement déposé sur la l'infrastructure d'hébergement de l'organisme demandeur ou fait l'objet d'un envoi chiffré.

C.2.2 Le SNDS



La base principale du Système national des données de santé (SNDS) regroupe et apparie les principales bases de données de santé nationales existant en France. C'est actuellement l'une des plus importantes bases de santé au monde : elle contient les informations de santé relatives à plus de 65 millions de Françaises et de Français.

Le SNDS base principale regroupe :

- les données de l'Assurance maladie (base SNIIRAM),
- les données des hospitalisations (base PMSI),
- les données des causes médicales de décès (base BCMD),
- les données relatives au handicap (en provenance des Maisons départementales des personnes handicapées - MDPH - données gérées par la Caisse nationale de solidaité pour l'autonomie - CNSA).

La gouvernance du SNDS est assurée par un pilotage ministériel et inclut les producteurs de données, parmi lesquels l'Inserm, fournisseur des données concernant les causes médicales de décès, via le CépiDc. Chaque producteur de données sources demeure responsable de ses traitements, la Cnam étant responsable du traitement de la base principale du SNDS.

C.2.3 L'accès au SNDS

Les conditions de la mise à disposition des données du SNDS sont définies par le code de la santé publique.

L'article L1461-1 III et R1461-1 du code de la santé publique définissent précisément les finalités de son utilisation.

Les finalités poursuivies doivent être d'intérêt public. Les données du SNDS peuvent ainsi être utilisées pour contribuer à la recherche, aux études, à l'évaluation et à l'innovation dans les domaines de la santé et de la prise en charge médico-sociale.

Il est en revanche interdit de procéder à un traitement des données dans les cas où celui-ci entrainerait :

- une prise de décision à l'encontre d'une personne physique identifiée sur le fondement des données la concernant et figurant sur l'un de ces traitements (condition qui ne s'applique pas aux données des causes de décès),
- la promotion des produits de santé en direction des professionnels de santé ou des établissements de santé,



- l'exclusion de garanties des contrats d'assurance ou la modification de cotisations ou de primes d'assurance pour un individu ou un groupe d'individus.

Les procédures d'accès aux données de santé reposent sur la distinction entre :

- les recherches impliquant la personne humaine (dite loi Jardé), qui concerne notamment les cas d'un appariement de données à celles issues d'un recueil impliquant la personne humaine,
- les recherches n'impliquant pas la personne humaine, qui concerne notamment la réutilisation de données déjà recueillies.

Pour les recherches impliquant la personne humaine (loi Jardé), le dossier doit être soumis au Comité de Protection des Personnes (CPP).

Pour les recherches n'impliquant pas la personne humaine, si le traitement n'est pas sous la responsabilité d'un organisme bénéficiant d'un accès permanent (cf. encadré), un dossier doit être soumis sur la plateforme du Health Data Hub qui dispose de 7 jours pour l'étudier et le transmettre :

- d'une part au Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la Santé (CESREES), qui se prononce sur sa conformité sous un mois,
- d'autre part à la Commission nationale informatique et libertés (CNIL), qui contrôle le niveau de sécurité des données et rend un avis dans un délai de 2 mois (renouvelable une fois).

L'accès permanent de l'Inserm au SNDS

L'Inserm, comme d'autres institutions publiques susceptibles de mettre en œuvre des traitements sur des données de santé, bénéficie d'un accès permanent aux données du Système national des données de santé (SNDS) pour les équipes et unités de recherche Inserm.

Dans ce cadre, le PDG de l'Inserm est chargé d'habiliter les utilisateurs de son organisme à accéder au portail et aux données du SNDS en tant qu'Autorité d'enregistrement (AE) pour l'Inserm. Ce pouvoir a été délégué à plusieurs Autorités d'enregistrement déléguées (AED). Ce sont ces derniers qui traitent les demandes d'accès aux données.

France Cohortes est en charge de la gestion des accès aux données du SNDS pour les personnels de l'Inserm dans le cadre de l'accès permanent.

C.2.4 La demande d'accès au SNDS

Pour accéder au SNDS, en l'absence d'un accès permanent, le dossier déposé sur la plateforme du HDH doit comprendre les documents suivants :



- un résumé de l'étude, recherche ou évaluation selon la grille retenue par le CESREES. Ce résumé doit impérativement être rédigé en français (5 pages maximum),
- un protocole scientifique incluant au moins les précisions demandées dans le résumé. Le protocole peut être soumis en anglais,
- la (es) déclaration(s) d'intérêt du(es) responsable(s) de traitement et du responsable de la mise en œuvre.
- la demande(s) d'autorisation CNIL,
- la lettre/notice d'information aux personnes concernées, de non opposition et/ou de consentement le cas échéant,
- la liste des financeurs de l'étude, le cas échéant,
- l'avis émis par le comité scientifique et/ou éthique qui a/ont évalué le projet et la composition de ce(s) comité(s), le cas échéant.

C.3 Les exploitations et travaux méthodologiques du CépiDc

C.3.1 Les analyses du CépiDc

Les analyses du CépiDc (cf.A.1.2, point b) ont porté sur les thématiques suivantes :

C.3.1.1 Les principales tendances de variation de la mortalité par cause depuis 2000

Cette étude⁵ visait l'évolution de la mortalité par cause en France métropolitaine entre 2000 et 2016. En 2016, pour 579 230 décès enregistrés, les effectifs de décès par tumeurs (29,0 %) et maladies cardiovasculaires (24,2 %) étaient prépondérants. La surmortalité masculine (sex-ratio de 1,7) était observée pour la quasi-totalité des causes. Dans un contexte de baisse générale de la mortalité, celle-ci a particulièrement diminué depuis 2000 pour le sida et le VIH, les accidents de transport, ainsi que pour les maladies cardiovasculaires pour les deux sexes. Elle a fortement augmenté pour le cancer du poumon chez la femme et de façon plus modérée pour les deux sexes pour le cancer du pancréas et le cancer du cerveau.

C.3.1.2 Les facteurs de variation de la mortalité par suicide

Une étude portant sur le **lien entre température ambiante et suicide**⁶, a mis en évidence une association positive sur l'ensemble du spectre des températures à la différence de la relation en

⁶ Lehmann F., Alary P.-E., Rey G., Slama R. Association of daily temperature with suicide mortality: A comparison with other causes of death and characterization of possible attenuation across five decades. Am J Epidemiol. 2022



⁵ Boulat T., Ghosn W., Morgand C., Falissard L., Roussel S., Rey G. Principales évolutions de la mortalité par cause sur la période 2000-2016 en France métropolitaine. Bull Epidémiol Hebd. 2019;(29-30):576-84.

U observée pour les autres causes de décès.

Une autre étude portait sur le **lien entre chômage et suicide** entre 2000 et 2010 à partir des données européennes⁷, et mettait en évidence une augmentation globale significative de 0,3 % du taux de suicide pour une augmentation de 10 % du taux de chômage. Cette association était significative dans trois pays : 0,7 % (intervalle de confiance à 95 % : [0,0 % ; 1,4%]) aux Pays-Bas, 1,0 % (intervalle de confiance à 95 % : [0,8 % ; 2,9 %]) en France.

D'autres travaux ont exploité les données du SNDS pour étudier les **contacts avec le système de soins des sujets décédés par suicide**⁸. Globalement, 8,5 % des suicidés ont consulté un médecin ou se sont présentés aux urgences le jour du décès, 34,1 % des suicidés la semaine précédant le décès, 60,9 % des suicidés le mois précédant le décès. La plupart des contacts concernaient un médecin généraliste ou une salle d'urgence. Au cours du mois précédant le suicide, 24,4 % des individus ont été hospitalisés au moins une fois. Les troubles mentaux (36,8 % des cas) étaient 7,9 fois plus fréquents chez les suicidés que dans la population générale. Les ratios les plus élevés estimés pour les maladies somatiques étaient pour les maladies du foie/pancréas (ratio standardisé de prévalence (RSP) = 3,3) et l'épilepsie (RSP = 2,7).

C.3.1.3 Les inégalités sociales et spatiales de mortalité par cause

Une analyse a estimé les indicateurs synthétiques d'inégalités sociales (SII) d'années de vie perdues par cause⁹ à partir des données de l'EDP chaînées aux causes médicales de décès. En moyenne, les hommes les moins éduqués ont perdu 6,6 (intervalle de confiance à 95 % = [5,9;7,3]) années de vie par rapport aux plus éduqués sur la période 1982-1990. Cette perte a augmenté au cours des deux périodes suivantes (1990-1996 et 1999-2005) pour atteindre 7,4 ([6,7;8,0]) années de vie. Les contributions les plus importantes au SII toutes causes confondues sur toutes les périodes pour les hommes sont les « autres » causes de décès (40 %) et la mortalité par cancer (30 %), tandis que les causes de décès externes et cardiovasculaires ont contribué entre 10 % et 20 %. Pour les femmes, la perte était stable sur toutes les périodes, avec un SII toutes causes confondues d'environ 3,5 années de vie perdues. En considérant la contribution de chaque cause de décès au SII toutes causes confondues, les « autres » causes de décès dominaient, suivies des causes cardiovasculaires à toutes les périodes.

Une autre analyse¹⁰, toujours à partir de l'EDP, a permis d'estimer simultanément l'association entre mortalité par causes et niveau socioéconomique individuel et collectif, à l'aide d'un modèle multiniveaux. Les résultats mettent en évidence une association positive et significative entre le désavantage social contextuel et la mortalité toutes causes confondues (Hazard

¹⁰ Ghosn W., Menvielle G., Rican S., Rey G. Associations of cause-specific mortality with area level deprivation and travel time to health care in France from 1990 to 2007, a multilevel analysis. BMC Public Health. 2017 Aug 2;18(1):86.



Dernière mise à jour le 11/08/2022

⁷ Laanani M., Ghosn W., Jougla E., Rey G. Association entre taux de chômage et suicide, par sexe et classe d'âge, en France métropolitaine, 2000-2010. Bul Epidemiol Hebd. 2015 Jan;1–2:2–6.

⁸ Laanani M., Imbaud C., Tuppin P., Poulalhon C., Jollant F., Coste J., Rey G., Contacts with Health Services During the Year Prior to Suicide Death and Prevalent Conditions: A Nationwide Study. J Affect Disord. 2020;23;274:174-182 9 Latouche A., Andersen P.-K., Rey G., Moreno-Betancur M. A Note on the Measurement of Socioeconomic Inequalities in Life Years Lost by Cause of Death. Epidemiology. 2019 Jul;30(4):569-572

Ratio (HR) = 1,24): les cancers, les maladies cérébrovasculaires, les cardiopathies ischémiques et les maladies évitables liées au soin et aux comportements à risque (HR de 1,14 à 1,29). Ces associations contextuelles ne présentent pas de différences substantielles selon le sexe, sauf pour la mortalité prématurée due aux cardiopathies ischémiques, beaucoup plus élevée chez les femmes. De manière inattendue, la mortalité diminuait à mesure que le temps nécessaire pour atteindre les structures de soins de santé augmentait. Seules les disparités géographiques de mortalité par maladies cérébrovasculaires et cardiaques ischémiques étaient fortement expliquées par des effets de composition et de contexte.

C.3.1.4 Les facteurs de variation de la mortalité infantile

Une étude¹¹ a porté sur les tendances récentes d'évolution de la mortalité infantile, uniquement à partir de données d'état civil. Elle a mis en évidence qu'entre 2000 et 2019, 24,4 % des décès sont survenus au cours du premier jour de vie et 47,8 % au cours de la période néonatale précoce. L'analyse Joinpoint a identifié deux points d'inflexion, en 2005 et 2012. Le taux de mortalité infantile a fortement diminué de 2001 à 2005 (pente : -0·0167 décès/1 000 naissances vivantes/mois ; intervalle de confiance à 95 % : [-0·0219 ; -0·0116]) puis diminué lentement entre 2005 et 2012 (pente : -0·0041 ; intervalle de confiance à 95 % : [-0·0065 ; -0·0016]). A partir de 2012, une augmentation significative du taux a été observée (pente : 0·0033; intervalle de confiance à 95 % : [0·0011 ; 0·0056]). Ces tendances étaient notamment induites par une augmentation de la mortalité pendant la période néonatale précoce.

Des travaux sur la distribution de la mortalité par Covid-19 devraient permettre d'analyser les inégalités sociales, spatiales, selon le pays de naissance et selon l'engorgement des services.

C.3.2 Les travaux méthodologiques du CépiDc

Les développements méthodologiques réalisés par CépiDc ont notamment porté :

- sur le **traitement automatique du texte médical**, pour contribuer à accroître la productivité de l'équipe de production de la base des causes médicales de décès. Une thèse¹² menée sur la mise en œuvre de méthode de *deep learning* sur les millions d'observations annotées que le CépiDc a produites a abouti à des résultats très prometteurs, permettant de porter des solutions innovantes pour les développements d'outils de codage automatiques envisagés par l'OMS et l'Iris Institute, susceptibles d'accélérer le traitement et d'accroître la comparabilité internationale des données. L'enthousiasme international suscité par ces excellents résultats a permis au CépiDc de proposer un projet de réplicabilité de ces méthodes sur les données des Etats-Unis, d'Angleterre, d'Italie et possiblement d'autres pays

- sur les techniques d'exploitation des données longitudinales, et notamment

¹² Louis Falissard. Deep learning methods in epidemiology and their applications to electronic health databases. Santé publique et épidémiologie. Université Paris-Saclay, 2021. English. (NNT : 2021UPASR019). (tel-03402715)



¹¹ Trinh N.T.H., de Visme S., Cohen J.-F., Bruckner T., Lelong N., Adnot P., Rozé J.-C., Blondel B., Goffinet F., Rey G., Ancel P.-Y., Zeitlin J., Chalumeau M. Recent historic increase of infant mortality in France: A time-series analysis, 2001 to 2019. Lancet Reg Health Eur. 2022 Mar 1;16:100339

l'introduction de méthodologies faisant appel au *machine learning* dans un objectif d'inférence causale. Ces méthodes constituent une clé de voute en vue d'une exploitation des données de cohortes et du SNDS à des fins de recherche épidémiologique.

Le CépiDc a pris à sa charge des **recherches portant sur la production ou les méthodes d'exploitation des données des causes médicales de décès**. Il a notamment réalisé :

- des travaux méthodologiques portant sur la redéfinition des indicateurs d'inégalités sociales et socio-spatiales de santé¹³,
- une revue de la littérature des méthodes de chaînage direct et indirect des données¹⁴,
- une étude de la sensibilité des analyses de variation à l'imputation des causes de décès manquantes¹⁵,
- une étude portant sur la problématique de la remontée d'information des certificats de décès par des instituts médico-légaux¹⁶,
- un développement méthodologique pour l'exploitation des causes multiples de décès¹⁸.

Des travaux sont en cours, notamment :

- avec le Centre international de recherche sur le cancer (CIRC): une étude vise à appliquer des méthodes de mesures de la survie nette, et à comparer les résultats obtenus avec des méthodes de mesures de risques par causes spécifiques sur les données de la cohorte EPIC,
- avec l'Assistance publique-hôpitaux de Paris (AP-HP) et l'Unité mixte de recherche HeKA (Inserm-Inria): élaboration d'algorithmes d'identification de diagnostics et codes CIM à partir des comptes rendus médicaux, par transfert de l'apprentissage effectué sur la base des causes médicales de décès,

¹⁸ Rey G, Piffaretti C, Rondet C, Lamarche-Vadel A, Moreno-Betancur M. Analyse de la mortalité par cause : pondération des causes multiples. Bul Epidémiol Hebd. 2017 Oct 1;(1):13–9.



¹³ Moreno-Betancur M, Latouche A, Menvielle G, Kunst AE, Rey G. Relative index of inequality and slope index of inequality: a structured regression framework for estimation. Epidemiology. 2015 Jul;26(4):518–27.

¹⁴ Bounebache SK, Quantin C, Benzenine E, Obozinski G, Rey G. Revue Bibliographique des Méthodes de Couplage des Bases de Données : Applications et Perspectives dans le Cas des Données de Santé Publique. Journal de la Societe Française de Statistique, 2018;159(3):79–123.

¹⁵ Moreno-Betancur M, Rey G, Latouche A. Direct likelihood inference and sensitivity analysis for competing risks regression with missing causes of failure. Biometrics. 2015 Jun;71(2):498–507.

¹⁶ Richaud-Eyraud E, Gigonzac V, Rondet C, Khireddine-Medouni I, Chan-Chee C, Chérié-Challine L, et al. État des lieux des pratiques et de la rédaction des certificats de décès par les instituts médicolégaux en France, en 2016, dans la perspective de la mise en place d'un volet complémentaire du certificat de décès. La Revue de Médecine Légale [Internet]. 2018 Feb 1 [cited 2018 May 14];9(1):1–9

¹⁷ Richaud-Eyraud E, Rondet C, Rey G. [Transmission of death certificates to CepiDc-Inserm related to suspicious deaths, in France, since 2000]. Rev Epidemiol Sante Publique. 2018 Mar;66(2):125–33.

 avec l'équipe Constances : développement et application des méthodes de positive and unlabeled learning et des terminologies médicales pour améliorer le phénotypage des maladies dans le SNDS.

Dans le cadre de projets plus anciens, les thématiques de la mortalité des migrants (projet MEHO 2009-2012), de la mortalité évitable (projet AMIEHS, 2008-2011) ont fait l'objet de travaux méthodologiques.

**

