

Dossier de demande de re- labellisation des séries issues de la statistique sur les causes de décès

Table des matières

1	Lettre d'envoi.....	5
2	Description de l'insertion et de l'organisation de l'unité chargée de la production statistique au sein de l'organisme.....	6
2.1	Inserm et CépiDc.....	6
2.2	Missions du CépiDc.....	6
2.3	Structuration en « unité de service ».....	7
2.4	Organisation de l'unité.....	8
2.5	Détermination des orientations en matière de statistiques, distinction statistiques officielles et communication.....	11
3	Place de l'organisme dans la production statistique du domaine, relations avec le Service statistique public.....	11
4	Séries pour lesquelles est demandée la labellisation d'intérêt général et de qualité statistique.....	19
5	Etat des lieux de la mise en œuvre des recommandations émises à la précédente labellisation.....	23
6	Description synthétique du processus de production « amont » des données administratives.....	29
6.1	Acte générateur et modèles de certificats.....	29
6.2	Certificat papier ou électronique.....	34
6.3	Circuit de remontée de l'information.....	34
6.4	Suivi de la réception des volets médicaux/contrôle de l'exhaustivité.....	36
	Suivi de la transmission des certificats papiers.....	36
	Suivi de la transmission des certificats électroniques.....	36
	Suivi de collecte et campagnes de relance.....	37
6.5	Saisie, numérisation, standardisation des certificats papiers et standardisation des certificats électroniques rejetés par le batch automatique de codage.....	38
7	Description des traitements statistiques « aval ».....	39
7.1	Présentation des bases intermédiaires.....	39
7.2	Présentation des traitements statistiques et de la qualité associée.....	39
	Contrôles et corrections.....	39

Exhaustivité – Appariement avec les données de l’Insee	40
Codification des causes médicales de décès.....	42
Géographie.....	47
Méthodes de calcul des indicateurs, corrections des variations saisonnières... ..	47
Données provisoires – <i>fast estimates</i>	47
Autres travaux de validation - évaluation de la campagne annuelle de production.....	48
7.3 Compléments méthodologiques relatifs à des enjeux ou séries spécifiques	49
8 Existence et gestion de ruptures de séries passées ou à venir.....	50
8.1 Evolutions passées et à venir	50
8.2 Processus de revue périodique des choix et spécifications méthodologiques.	52
9 Gouvernance et rapports d’audit récents.....	52
9.1 Mission interIG sur les statistiques de décès et mission de rattrapage	52
9.2 Gouvernance.....	53
10 Lettre d’engagement sur le calendrier de publication et de diffusion des séries	54
11 Modalités de présentation et de diffusion des séries	54
11.1 Accessibilité.....	54
11.2 Identification des séries labellisées	55
11.3 Calendrier prévisionnel de publication.....	55
11.4 Politique de correction d’erreur	55
11.5 Explication des sigles et concepts.....	55
11.6 Description des nomenclatures de diffusion	55
11.7 Documentation articulée selon différents niveaux	56
11.8 Liens vers des sites partenaires	57
11.9 Modalités d’accès aux données pour les chercheurs.	57
12 Gestion de la confidentialité statistique.....	57
13 Moyens humains et financiers mis en œuvre pour les différentes étapes du processus statistique	58

14	Autres projets et objectifs des prochaines années pouvant influencer sur les séries publiées de façon indirecte	60
	Fondements juridiques	61
	Références	63
	Liste des annexes	66
	Liste des annexes sous forme de tableaux	67

1 Lettre d'envoi

Lettre d'envoi adressée à la Présidente de l'Autorité de la statistique publique avec copie à la Présidente du Comité du label de la statistique publique. Cette lettre engage l'organisme dans son ensemble sur le processus de labellisation. Elle sera jointe au dossier de présentation – vérifier statut note Inserm et engagement

Copie d'écran ci-dessous - Aussi en Annexe 1 - A1-2024-036_ASP_M Elbaum_Dossier soumission comité label statistique publique envoyée le 21 novembre 2024.



ANNEXE 1

Le Président-directeur général

Dossier suivi par :

Elise Coudin

Directrice

Centre d'épidémiologie des causes de décès

elise.coudin@inserm.fr

N°réf. DG-2024-036

Madame Mireille Elbaum

Présidente de l'Autorité de la statistique publique

88 avenue Verdier

CS 70058

92541 Montrouge Cedex

Paris, le 21 novembre 2024

Objet : Dossier technique pour soumission au comité du label de la statistique publique

Madame la Présidente,

Par le présent courrier et faisant suite au délibéré sur les statistiques de causes de décès du 21 février 2024, je vous confirme officiellement la volonté de l'Inserm de fournir à l'Autorité de la statistique publique, d'ici la fin de l'année 2024, un dossier « technique » de re-labellisation des statistiques des causes de décès afin qu'il puisse commencer à être instruit par le Comité du label.

En 2024, la modernisation de la production de la statistique sur les causes de décès s'est poursuivie dans le cadre du Centre d'épidémiologie sur les causes médicales de décès (CépiDc), unité de service relevant de l'Inserm. Cette modernisation s'inscrit dans le cadre des exigences du code des bonnes pratiques en matière de statistiques officielles. Les données sur les causes de décès en 2022 ont été diffusées à Eurostat fin septembre, soit trois mois avant la fin du délai réglementaire. Elles ont été accompagnées de publications conjointes Inserm-CépiDc, Santé publique France et Direction de la recherche, des études, de l'évaluation et des statistiques (DREES). La Direction générale de la Santé, la Direction générale de la recherche et de l'innovation, la DREES et l'Institut national de la statistique et des études économiques, ont demandé à l'Inserm de mettre en place une gouvernance pérenne de la production de cette statistique, instituant, avec l'Inserm, un comité stratégique de la statistique sur les causes de médicales de décès, qui reprend le format de l'ancien comité de pilotage de la mission de rattrapage et de rénovation. Ce comité s'est réuni en octobre pour la première avec ce nouveau rôle.

L'Inserm ne perd pas de vue que, selon les termes de votre délibéré, la re-labellisation ne pourra être accordée qu'une fois que les conditions de ressources financières et humaines, et les ressources informatiques adaptées soient garanties. Il est engagé dans un dialogue actif avec ses deux tutelles ministérielles pour l'obtention des moyens nécessaires à la satisfaction de cette exigence.

Je vous prie d'agréer, Madame la Présidente, l'expression de mes salutations distinguées.

A handwritten signature in blue ink, appearing to read 'D. Samuel'.

Pr Didier Samuel
PDG de l'Inserm

2 Description de l'insertion et de l'organisation de l'unité chargée de la production statistique au sein de l'organisme

Statut et mission de l'organisme, service responsable des statistiques, organigramme et compétences des équipes statistiques, détermination des orientations en matière de statistiques, distinction claire entre les publications statistiques et la communication (se rapporte au principe n° 1 du CBPSE).

2.1 Inserm et CépiDc

L'Inserm est un opérateur public, organisme de recherche sous le statut d'EPST sous la double tutelle du ministère en charge de la santé (DGS) et du ministère en charge de la recherche (DGRI). Le CépiDc, Centre d'épidémiologie sur les causes médicales de décès, est l'unité de service de l'Inserm qui a la mission de production et diffusion de la statistique sur les causes de décès.

Au sein de l'Inserm, l'activité du CépiDc est dans le périmètre de l'Institut Thématique de Santé Publique (ISP). L'évaluation professionnelle de la directrice du CépiDc est assurée par le directeur de l'ISP. L'ISP participe à l'arbitrage des demandes de budget et de postes. Le CépiDc est sous la responsabilité administrative de la Délégation Régionale Paris 11 de l'Inserm. Ceci se matérialise par la prise en charge de la gestion de l'hébergement et de toutes les formalités administratives et financières, et par une participation à l'arbitrage des demandes de budget et de postes.

Le CépiDc s'appuie sur les services supports de l'Inserm. Le Département Système d'Information de l'Inserm assure, en propre ou par le biais de prestataires externes, l'hébergement, la maintenance et la mise en œuvre des développements du système d'information. Le Département des Affaires Juridiques valide la mise en place des conventions de partenariats.

2.2 Missions du CépiDc

Le CépiDc, Centre d'épidémiologie sur les causes médicales de décès, a pour mission de produire la base de données statistique sur les causes médicales de décès en France, de la diffuser et de réaliser des analyses sur cette base de données. Cette base de données statistique repose sur la collecte et le traitement des volets médicaux des certificats de décès. Cette mission est encadrée par la loi nationale ([Articles L2223-42](#) et [R2213-1-1](#) du code des collectivités territoriales) qui assure le rôle de l'Inserm dans la collecte de l'information et précise les finalités d'usage des données. Ainsi, au niveau national, les données collectées et traitées par le CépiDc sont mobilisées à des fins de veille sanitaire, de statistiques

publiques et de recherche uniquement : [Articles L2223-42](#) : « Ces informations ne peuvent être utilisées que pour des motifs de santé publique :

1° A des fins de veille et d'alerte, par l'Etat, les agences régionales de santé et l'Agence nationale de santé publique ;

2° Pour l'établissement de la statistique nationale des causes de décès et pour la recherche en santé publique par l'Institut national de la santé et de la recherche médicale ;

3° Pour les traitements de données concernant la santé, dans les conditions fixées à l'article [L. 1461-3](#) du code de la santé publique ;

4° Pour alimenter le système national des données de santé défini à l'article [L. 1461-1](#) du même code ;

5° Pour l'établissement de statistiques dans le cadre de l'[article 7 bis de la loi n° 51-711 du 7 juin 1951](#) sur l'obligation, la coordination et le secret en matière de statistiques, par l'Institut national de la statistique et des études économiques ou par les services statistiques du ministre chargé de la santé. Ces données doivent être conservées séparément des données du répertoire national d'identification des personnes physiques détenues par l'Institut national de la statistique et des études économiques. »

Au niveau international, la statistique sur les causes de décès est soumise au règlement [CE 1338/2008](#) relatif aux statistiques communautaires de la santé publique et de la santé et de la sécurité au travail et au règlement d'application [UE 328/2011](#). Selon ces règlements, la collecte doit suivre les recommandations de l'Organisation mondiale de la Santé (OMS). La statistique doit être codée dans la classification internationale des maladies de l'OMS ; les normes d'évaluation de la qualité sont celles du code des bonnes pratiques en matière de statistique européenne ; les concepts, champs, variables, périodes de référence et délais de transmission sont fixés. De ce fait, le CépiDc est considéré par l'Insee comme une « autorité statistique nationale », productrice de statistique officielle, hors service statistique public, auquel s'applique le règlement UE 223, et le code des bonnes pratiques en matière de statistique européenne.

Le CépiDc est membre du Centre Collaborateur français de l'OMS pour la Famille des Classifications Internationales (FIC) de laquelle la classification internationale des maladies découle, et membre de l'Iris Core Group, groupe de pays en charge de maintenir et développer le logiciel international de codage automatique des causes médicales de décès IRIS.

2.3 Structuration en « unité de service »

Le CépiDc est structuré actuellement en « unité de service » à durée limitée de 5 ans (et ce depuis 2011). Cette unité de service prend fin en 2025. Elle entre dans le cadre de la campagne de demande de renouvellement des unités de service qui s'accompagne d'une évaluation de la structure en interne à l'Inserm.

Les unités de service font partie des « autres formations de recherche et d'appui à la recherche ». Les textes réglementaires portant sur l'organisation de l'Inserm ([articles R324-1 à 3](#) du code de la recherche) indiquent :

« Les autres formations de recherche ou d'appui à la recherche sont créées, modifiées et supprimées par décision du président de l'institut, le cas échéant conjointement avec l'autorité compétente de l'organisme partenaire.

Les responsables des formations de recherche ou d'appui à la recherche sont désignés par décision du président de l'institut, le cas échéant conjointement avec l'autorité compétente de l'organisme partenaire. »

Selon la [décision DAJ2013-110](#) du PDG de l'Inserm sur les typologies des formations de recherche, une unité de service est créée :

« afin de mutualiser des moyens d'appui à la recherche ou d'exercer des activités concourant à la recherche. Son objet peut être scientifique, technologique ou de gestion. »

Les unités de service sont d'une durée limitée, précisée dans la décision de création (mais renouvelables). Les formations de recherche (dont les unités de service) ont en charge :

« la réalisation des missions confiées à l'organisme (l'Inserm) :

- Développement de la recherche*
- Contribution aux actions de transfert dans les domaines de la valorisation économique et sociale, de la formation à la recherche et par la recherche, de l'information et de la coopération internationale. »*

Pour autant, le CépiDc ne remplit pas proprement dit une « mission d'appui à la recherche mutualisable au sein de l'Inserm ». Il remplit une mission confiée à l'Inserm mais qui n'apparaît pas directement dans les textes relatifs à l'organisation de l'Inserm ([articles R324-1 à 3](#), code de la recherche et [décision DAJ2013-110](#)).

La structuration du CépiDc en unité de service a des implications qui doivent se lire sous le prisme du code des pratiques.

L'avantage principal est que le cadre réglementaire des unités de service de l'Inserm est plus souple et donc plus aisément adaptable au cas d'un ONA que celui des unités de recherche, notamment en termes d'évaluation, et de désignation du directeur, à condition que les critères d'évaluation/accompagnement appliqués, et plus largement le cadre de fonctionnement soient cohérents avec le code des pratiques (profil et mode de recrutement du directeur, visibilité sur les moyens, articulation avec les missions complémentaires). Dans le cadre du renouvellement de l'unité en 2025 et suite à la demande du comité stratégique qui assure la gouvernance de la statistique sur les causes de décès (voir après), il a été acté qu'un cadre supérieur de la statistique publique participerait au comité venant visiter le service (visite programmée courant avril). Ce comité émet des recommandations adressées au PDG de l'Inserm qui permettent à celui-ci d'orienter la décision de renouvellement de l'unité. L'Inserm va lancer une réflexion en interne pour voir si ces implications justifient de faire évoluer le cadre lors de la demande de renouvellement de l'unité.

L'unité de service est dotée d'un **règlement intérieur** (Annexe). Celui-ci précise les missions du service, les conditions dans lesquelles ces missions sont accomplies, la composition de la gouvernance, les missions de statistique publique ainsi que le cadre réglementaire et de qualité dans lequel elles s'inscrivent.

2.4 Organisation de l'unité

- *Direction* - une directrice, statisticienne (Elise Coudin, inspectrice générale de l'Insee, PhD en économétrie, détachée de l'Insee) et une assistante-gestionnaire chargée de communication (assistante ingénieure de l'Inserm).

L'unité comprend deux pôles :

- le *pôle production* (responsable Diane Martin, ingénieure de recherche de l'Inserm, pharmacienne, adjoint Pierre Boulet, ingénieur d'études, statisticien démographe) a pour mission de réaliser l'ensemble des traitements relatifs au recueil, au codage, à la constitution de la base par traitements statistiques et à sa validation – il est composé :

- d'une équipe automatisation qui tient le système-expert de codage automatique, les modèles de réseaux de neurones en production, en développement, les traitements statistiques amont et aval jusqu'à la constitution des bases de données, les flux d'entrée et de sortie de l'information entre le CépiDc et les partenaires et la veille technologique. Cette équipe est composée d'une responsable en traitement automatique (ingénieure d'étude sénior de formation informatique spécialisée en traitement du langage, responsable), d'un datascientist (ingénieur de recherche senior, formation Bac+5 en économie puis certificat en sciences des données), d'une dataengineer (ingénieure d'études junior, formation BAC+ 5 en sciences des données), d'un informaticien – responsable d'application en cours de recrutement (M2).

- d'une équipe d'experts-nosologistes du codage. La responsable de l'équipe de codage, la coordonnatrice de proximité et l'expert en charge de la CIM11 ont chacun une formation médicale et/ou en santé publique – sage-femme+ M2 de santé publique, infirmière, M2 en santé publique + études de médecine) et sont en charge de la participation à l'évolution et à la maintenance de la classification internationale des maladies sur les aspects mortalité dans le cadre du centre collaborateur français à l'OMS, de l'organisation du codage dans l'équipe de codage et de la formation, du codage des cas les plus complexes, de l'enrichissement du dictionnaire du système de règles permettant le codage automatique par batch. Trois experts-codeurs en plus (ingénieurs d'études, infirmiers), trois nosologistes (assistants ingénieurs, formations de technicien d'information médicale ou formation interne) et deux codeuses (techniciennes, formations internes) assurent le codage interactif des certificats. Il y a aussi recours à des prestations externes de codage pour l'équivalent de 0,8ETP. Une assistante ingénieure en plus est chargée du suivi de la collecte, relance des ARS et du prestataire pour limiter les trous de collecte.

- le *pôle diffusion* a pour mission de participer à la validation de la base de données, d'assurer sa diffusion sous divers formats (agrégés, détails, opendata, à la demande) et d'en réaliser une exploitation pour des analyses statistiques ou épidémiologiques. Il veille à ce que ces diffusions respectent les règles de confidentialité/secret et met en œuvre les méthodes appropriées pour cela. Il peut être amené à mettre en place et exécuter des conventions de mise à disposition de données pour certains projets de recherche nécessitant des données plus fines et indispensables pour les appariements que celles accessibles au SNDS. Il est composé de deux agents : une statisticienne responsable, Fanny Godet (administratrice de l'Insee statisticienne, Ph.D en mathématiques, en détachement de l'Insee), et un data-analyst (M2 de sociologie quantitative).

L'équipe au total compte 21 agents + 1 en cours de recrutement. L'organigramme fonctionnel reporté ci-dessous indique les postes financés par le Plan de rattrapage et de refonte (dont ceux demandés à être pérennisés en jaune) en plus de ceux entrant déjà dans la trajectoire d'emploi de l'Inserm (15 supports Inserm ou Inserm). L'Inserm porte en effet une demande de 5 postes supplémentaires (2 d'investissement

et trois de structure) à partir de 2026, aux 15 postes déjà gagés par l'Inserm auprès de ses tutelles.

Organigramme fonctionnel au 23/01/2025			
Pôle direction	ETP	Fonctions	statut
directrice	1	pilotage de l'unité : orientation stratégique de l'unité, stratégie de production, diffusion, pilotage scientifique et administratif en lien avec l'encadrement réglementaire et la gouvernance..., compétences statistiques, managériales et stratégiques avérées (cf code bonnes pratiques statistique européenne) : cadre supérieur de la statistique publique senior ou équivalent	IG insee détachement - support Inserm
assistante - gestionnaire - chargée de communication	1	communication interne et externe, gestion des ressources/dépenses, secrétariat	AI Inserm
Pôle diffusion			
responsable - statisticienne (IR)	1	diffuser, valoriser et réaliser des analyses statistiques et épidémiologiques sur les causes de décès ; validation des données; profil statisticien senior(administrateur Insee ou ingénieur de recherche encadrant) ; avec expérience en diffusion (secret; dataviz; études)	admin Insee détachement - support Inserm
data analyst (IR)	1	diffuser, valoriser et participer à des analyses statistiques et épidémiologiques ; extraction de données; profil statisticien junior	CDD - support Inserm
Pôle production			
responsable de pôle - IR	1	piloter et mettre en œuvre la stratégie de production lors des campagnes de production; approche qualité ; ingénieur de recherche senior, encadrement	IR Inserm
adjoint à la resp de pôle - statisticien IE	1	statisticien, traitements statistiques sur la base de donnée, pilotage et suivi de la collecte, encadrant en l'absence du responsable du pôle	dispo - CDD Projet
chargée qualité et suivi collecte AI	1	suivi de collecte, édition des variables non médicales	AI - Inserm
Equipe automatisé			
responsable équipe - ingénieure TAL IE	1	pilotage des étapes d'automatisation, participation au développement du logiciel système expert Iris/Muse international ; compétences en TAL pour optimiser le codage automatique ; supervision du SI	IE Inserm
datascientist IR	1	datascience, développement de modèles de deep learning, R&D et mise en production, investissement sur transition CIM 11	dispo - CDD Projet
dataengineer IE	1	datascience, développement de modèles de deep learning, R&D et mise en production	CDD Projet 2025- Inserm 2026
informaticien -responsable d'application (IE)	(vac)	responsable de l'application Iris/Muse, mise en œuvre et suivi de production; traitements informatiques relatifs et spécification des évolutions du SI	CDD Projet
Equipe codage			
responsable équipe - experte codeuse IE	1	responsable, référente mortalité pour la France à l'OMS, organisation de l'équipe de codage, formation et codage expert ++, connaissances CIM et médicales, niveau IE senior + responsabilités d'encadrement	IE Inserm
coordinatrice - experte codeuse IE	1	organisation pratique de l'équipe de codage, formation et codage expert, maintien des tables du dictionnaire, niveau IE senior + responsabilités d'encadrement de proximité	CDD - support Inserm
experts codeurs IE	4	codage expert, connaissances médicales requises, codage expert et expertise dans la CIM 11, maintien des tables du dictionnaire, connaissances médicales requises, niveau IE	CDD - Projet (2+2)
Noso codeurs AI	3	codage manuel classique, connaissances CIM	3 AI - Inserm
codeur T	2	codage manuel classique	2T - Inserm
+ recours à prestation codage			
Total ETP - 21 - 22			

2.5 Détermination des orientations en matière de statistiques, distinction statistiques officielles et communication

détermination des orientations en matière de statistiques, distinction claire entre les publications statistiques et la communication (se rapporte au principe n° 1 du CBPSE).

La statistique étant soumise aux règlements statistiques communautaires, les orientations en matière de statistiques suivent le code des pratiques et sont proposées par le CépiDc au comité stratégique sur la statistique des causes de décès qui assure la gouvernance de la statistique sur les causes de décès (voir paragraphe dédié à la gouvernance). Cette gouvernance et les principes du code des pratiques sont inscrits dans le règlement intérieur de l'unité fourni en annexe.

Par exemple, le comité stratégique sur la statistique des causes de décès a acté en février 2025 le calendrier de diffusion des données 2023 et suivantes : « Le comité stratégique acte que le calendrier de diffusion des causes de décès données finalisées ne peut pas être réduit à moins de T+16-17 mois dans les prochaines années. Ce point pourra être révisé lorsque la couverture de la certification électronique sera bien plus importante et/ou dans le cas d'une profonde refonte de l'étape de synchronisation. Le comité stratégique note que des estimations provisoires des causes de décès de l'année N pourront en revanche être disponibles à T+6 mois, sur la base d'estimations pour corriger des biais induits par les trous de collecte et avant synchronisation avec les décès Insee, comme cela s'est déjà fait à T+12 mois puis T+9 mois les deux dernières années.[...]. Les notes préparatrices et comptes rendus des deux dernières réunions du comité stratégique sont fournis en annexe.

3 Place de l'organisme dans la production statistique du domaine, relations avec le Service statistique public

3.1 – Statistique de référence en santé publique, épidémiologie, démographie et statistique

Couverture sectorielle/populationnelle de l'organisme (et part dans son secteur/la population),

Les données de causes de mortalité sont une statistique de référence en santé publique, épidémiologie, démographie et statistique depuis la fin du XIX^e siècle. Elles sont mobilisées dans le cadre de la veille sanitaire, des politiques de santé publique et dans ceux de la statistique et de la recherche en santé.

La collecte des causes de décès et la classification internationale des maladies se sont construites de concert au cours du XIX^e-XX^e siècle en France et à l'international (voir [Coudin, Robert 2024](#)). Collecte et codification suivent les standards internationaux de l'Organisation mondiale de la santé pour assurer la comparabilité dans le temps et dans l'espace des données et des analyses qui en découlent. Cette conformité est obligatoire en Europe, car cette statistique fait partie des statistiques communautaires de la santé publique. La collecte de cette statistique s'articule à celle de l'Etat civil. Comme dans de nombreux pays, chaque décès sur le territoire donne lieu à la rédaction d'un certificat par un médecin ou professionnel de santé, dont la validation est nécessaire pour fermer le cercueil et procéder à l'inhumation. La forme du certificat de décès est fixée dans le Code général des collectivités territoriales. Il retranscrit nationalement les obligations internationales en la matière. Le volet administratif du certificat de décès avec les noms et les prénoms est destiné aux opérateurs funéraires et à la mairie qui dressera l'acte de décès de l'état civil et remplira le BEC associé à destination de l'Insee. Le volet médical,

anonyme, est transmis à l'Inserm. Conforme au format standard international décrit dans la classification internationale des maladies (CIM) tenue par l'OMS, ce volet médical recueille l'avis du médecin certificateur sur les maladies et les conditions impliquées dans le décès, leur enchaînement, et leur place dans le processus morbide conduisant au décès.

Alors que dans la plupart des pays européens, la mission de production et de diffusion de la Statistique sur les causes de décès est assurée par l'institut national statistique (INE, Istat, CBS, Destatis, ONS...), en France, elle est confiée à l'Inserm-CépiDc. En effet, en 1968, à l'occasion d'une nouvelle révision de la CIM, la mission de codification des causes de décès est confiée à l'Inserm afin qu'elle soit réalisée par un personnel médical qualifié et de manière centralisée. Jusqu'en 1987, les professionnels médicaux de la Section information en santé publique de l'Inserm, puis du Service commun d'information sur les causes médicales de décès (SC8) codifient des causes et reportent les codes sur les bulletins de décès compilés ensuite par l'Insee. À partir de 1988, avec le développement de l'informatisation, l'Inserm obtient la maîtrise complète du processus de production jusqu'à la diffusion : l'Insee met à disposition un extrait informatique de son fichier démographique de mortalité. Les experts-nosologues y adossent les codes des causes et l'Inserm peut ainsi publier des statistiques provisoires, puis définitives. Depuis 1997, le Centre d'épidémiologie sur les causes médicales de décès de l'Inserm (CépiDc, remplaçant le SC8 chargé du codage et de l'exploitation et le SC25 chargé de la saisie) assure la production et la diffusion de la statistique.

À partir de 2007-2008, avec la mise en place du système statistique européen, la Statistique sur les causes de décès devient une statistique officielle européenne soumise au règlement 223. Elle entre dans le champ d'application du règlement relatif aux statistiques communautaires de la santé publique ainsi que de la santé et de la sécurité au travail. Le règlement CE 1338/2008 et son règlement d'application UE 328/2011 (voir annexe et références juridiques) assurent la comparabilité de la statistique au sein de l'Europe et définissent les conditions de collecte de l'information. Celle-ci s'appuie sur des certificats de décès nationaux conformes aux recommandations de l'OMS et sur un codage des causes dans la classification internationale des maladies de l'OMS. Les normes d'évaluation de la qualité sont celles du code des bonnes pratiques en matière de statistique européenne (pertinence, exactitude, actualité, ponctualité, accessibilité, comparabilité, cohérence) ; les concepts, les champs, les variables, leur ventilation, les périodes de référence et les délais de transmission sont fixés. En particulier, les États membres doivent transmettre à Eurostat les données d'une année dans un délai de 24 mois à compter de la fin de cette même année. Le CépiDc de l'Inserm devient en 2017 une « autorité statistique nationale » (ONA : Other National Authority) productrice de statistique officielle, hors service statistique public (voir Annexe Lettre du DG de l'Insee au PDG de l'Inserm sur le statut d'ONA).

3.2 Partenariats

Relations et partenariats sur le champ statistique avec le SSP ou d'autres acteurs, transmission du programme statistique au Conseil National de l'Information Statistique (Cnis) et participation à ses réunions / travaux.

Le CépiDc de l'Inserm se situe dans un éco-système riche : statistique publique et de santé publique (OMS) au niveau international, veille sanitaire, état-civil, statistique publique et recherche en santé au niveau national. On détaille par la suite les grands partenariats suivant leur place dans le processus de production statistique. Pour les partenariats dans le cadre de la gouvernance, voir la section dédiée. Le programme de travail du CépiDc n'est pour le moment pas transmis / présenté au Cnis.

Partenariats dans la collecte de l'information

- L'information (certificats de décès, dont partie médicale) est collectée auprès des médecins certificateurs (et les infirmiers le cas échéant). Leur activité s'inscrit dans un cadre institutionnel variable (ville, hôpital, maison de retraite...).

- La Direction Générale de la Santé (DGS) est en charge du suivi et de l'évolution du cadre légal et opérationnel dans lequel s'inscrit la certification des décès, notamment le format du certificat de décès (décrets, arrêtés) et les modalités de transmission des informations en particulier à l'Inserm. Une convention lie la DGS et l'Inserm pour définir les responsabilités respectives dans la mise en œuvre de la *certification électronique*. En pratique la DGS est maîtresse d'ouvrage de la certification électronique, l'Inserm-CépiDc maîtresse d'ouvrage pour le volet médical et le DSI de l'Inserm est maîtresse d'œuvre de l'application de certification électronique CertDc. Le partenariat CépiDc-DGS est fort en ce moment (la DGS est représentée dans la gouvernance du CépiDc) notamment pour accompagner le déploiement de la certification électronique, et l'expérimentation d'ouvrir la possibilité de certification aux infirmiers.
- Les communes dont le service d'état civil est chargé de réceptionner le certificat de décès survenus sur leur territoire sous format papier, d'établir l'acte de décès, et d'émettre un bulletin d'état civil nominatif (B7bis) à l'attention de l'Insee pour mise à jour du RNIPP et l'établissement des statistiques démographiques, et une version sans les noms et prénoms (B7) pour accompagner le volet médical, lesquels sont envoyés à l'Agence Régionale de Santé (délégation départementale). On peut considérer qu'il y a un partenariat, même si le CépiDc n'intervient jamais directement auprès des communes (trop nombreuses) mais via l'intermédiaire des ARS locales.
- Les Agences Régionales de Santé (ARS) qui, veillent à la remontée des volets médicaux (+B7) depuis les communes de décès, les ouvrent et les transmettent à l'Inserm. Elles participent aussi à la veille sanitaire avec Santé publique France (voir après). Le CépiDc intervient auprès des ARS dans des objectifs d'amélioration de la collecte (certificats papiers), de relais auprès des mairies. Ces deux éléments sont parfois ressentis comme une charge sans plus-value par les ARS.

Partenariats dans la production après collecte

Insee – L'Insee (département de la démographie, répertoires des personnes, pôle répertoires et fichiers démographiques, et en fin de processus la division enquêtes et études démographiques de l'unité des études démographiques et sociales) est le partenaire essentiel à la production de la base, dans l'étape de gestion de l'exhaustivité de la collecte (dédoublonnage, gestion de la non-réponse totale), appelée « synchronisation ». Cette étape permet aussi des corrections et des enrichissements de certaines caractéristiques socio-démographiques de la base des causes médicales de décès. Le règlement statistique démographique et celui sur les causes de décès réfèrent au même champ / période et les réponses des Etats doivent être en cohérence (mêmes nombres de décès ventilés). Les informations sur le décès enregistré à l'Etat civil et sur le volet médical sont issues du même événement générateur : un médecin constatant le décès. Pourtant les circuits de remontée de l'information différents et la non-transmission à l'Inserm des noms et prénoms des personnes décédées, induisent des écarts, des délais et justifient cette approche de synchronisation détaillée par la suite.

Ce partenariat est régi par une convention relative à l'échange de données individuelles sur les décès entre l'Inserm et l'Insee n°2021122NF, des comités de suivi trimestriels entre l'Insee et le CépiDc.

DREES – Dans le cadre du projet de rattrapage sur les causes de décès, la DREES (directeur de projet + chargé d'étude sur les causes de décès) a travaillé de concert avec le CépiDc pour la mise en place d'une méthodologie mobilisant des neurones profonds classifiant les causes de décès en particulier pour décès 2018 et 2019. Depuis fin 2024, il n'y a plus d'échanges méthodologiques ou d'évaluation de l'approche mise en production au-delà des échanges ayant lieu lors des réunions de gouvernance, le CépiDc ayant les compétences en interne et le rattrapage étant terminé. En revanche, la DREES accompagne le CépiDc sur les aspects informatiques (suivi en production et projet de refonte informatique applicative). La DREES met aussi à disposition du CépiDc une bulle au CASD qui est mobilisée pour entraîner les modèles d'IA, voire faire de l'inférence, en développement et si besoin pour la production.

Homologues internationaux dans le cadre de l'Iris Core Group – Le CépiDc de l'Inserm fait partie de l'Iris Core Group de [l'Iris Institute](#), qui développe et maintient le logiciel Iris- Muse, principal outil (système expert de règle) de codification des causes de décès dans la classification internationale des maladies. L'institut Iris est hébergé par le Federal Institute for Drugs and Medical Devices (Bfarm) en Allemagne. L'Iris Core Group comprend des producteurs de la statistique des causes de décès, principalement des INS (Istat, CBS, ONS, Destatis, KSH hongrois) ainsi que le National Center for Health Statistics de la CDC pour les Etats-Unis, Bfarm et CépiDc. Le logiciel Iris est construit de façon à séparer tout ce qui relève de la langue de certification de tout ce qui relève des causalités, liens permettant la classification et la détermination d'indicateurs statistiques comme la cause initiale, ce qui permet d'utiliser le logiciel dans des pays ne parlant pas la même langue et d'assurer la comparabilité régionale de la statistique. Au-delà du groupe qui maintient le logiciel, ce dernier est mis à disposition gratuitement et est utilisé dans la grande majorité des pays européens. L'Iris Core Group se réunit deux fois une semaine par an pour valider les évolutions du logiciel, la stratégie dont son financement.

OMS – Le CépiDc est membre du Mortality Reference Group (groupe de référence sur la mortalité), le groupe de l'Organisation mondiale de la santé (OMS) chargé du maintien et des évolutions des règles de codage de la CIM en matière de mortalité, ainsi que d'autres sous-groupes qui en découlent (digitalization rules, review of rules... à noter que dans le cadre de la CIM 11, laquelle est fondamentalement numérisée, l'OMS développe en propre des outils de codage automatique). Le travail en collaboration ([Family of international classification network](#)) avec l'OMS est canalisé au niveau national par le centre collaborateur français de l'OMS (CCOMS) sur la famille des classifications internationales présidé par l'Agence du numérique en santé. Ce centre regroupe l'Inserm-CépiDc (pour la mortalité), la Caisse nationale de l'assurance maladie (qui intervient sur la nomenclature de classement des remboursements, appelée l'International Classification of Health Interventions (ICHI), l'Agence technique de l'information sur l'hospitalisation (pour la Classification internationale des maladies et ICHI en morbidité), l'École des hautes études en santé publique (pour la Classification internationale du fonctionnement, du handicap et de la santé). Les réunions des groupes à l'OMS sont régulières (deux semaines de réunions officielles par an + des réunions virtuelles trimestrielles pour certains sous-groupes). Celles du CCOMS au niveau français sont trimestrielles.

Eurostat – Il n'y avait plus de working group régulier concernant la statistique sur les causes de décès. Le sujet est généralement abordé dans le cadre du Working Group on Public Health – administrative data auquel assiste la DREES (et pas le CépiDc), puis au niveau des réunions des directeurs des statistiques sociales. Pour accompagner les pays au changement de nomenclature (passage à la CIM 11) et pour adapter collectivement les indicateurs statistiques d'Eurostat (agrégations de causes en short list européenne, principales ventilations de causes des statistiques diffusées), Eurostat a lancé depuis 2023 une Task Force ICD11 à laquelle participe le CépiDc pour la France aux côtés des autres états membres. Il y a deux réunions par an, des enquêtes régulières permettant de faire état des avancées des différents pays dans la préparation du changement et du travail collectif de mise à jour des indicateurs.

Partenariats dans l'exploitation première et la première diffusion

Veille sanitaire – Santé publique France – Les données brutes sur les causes de décès sont dès leur intégration au CépiDc mises à disposition en temps réel à Santé publique France à des fins de veille sanitaire. Elles sont mobilisées dans les bulletins hebdomadaires de mortalité, ainsi que ceux des suivis des épidémies, aux côtés d'autres sources. Les données issues de la certification électronique reflètent le temps réel, celles issues de la certification papier sont aussi partagées au moment de leur intégration au CépiDc. Un comité mensuel de suivi entre le CépiDc et Santé publique France a lieu. Ce partenariat est régi par une convention de collaboration entre Santé publique France et le CépiDc de l'Inserm qui couvre aussi des collaborations d'études et de recherche.

ARS Ile-de-France – une expérimentation actuellement (sous convention) conduit à dupliquer le flux de données envoyé à Santé publique France avec l'ARS Ile de France dans le cadre de développement d'indicateurs complémentaires de veille sanitaire plus localisés.

La publication marronnière accompagnant la sortie nationale de la statistique annuelle est depuis fin 2022 réalisée conjointement entre le CépiDc, **Santé publique France** et la **DREES**. Deux publications – un article dans le [bulletin épidémiologique hebdomadaire](#) et un [Etudes et résultats](#) – sortent en même temps que la diffusion de la statistique sur le site du CépiDc.

La base de données détail est une des sources principales du [système national des données de santé](#). Le CépiDc la fournit à la Cnam après diffusion nationale et internationale de la statistique. A ce titre, le CépiDc participe au comité de suivi du SNDS animé par la DREES (AMDAC) entre producteurs de la base principale du SNDS (Cnam, ATIH, DREES, HDH, Cnsa).

Partenariats dans les études, la méthodologie, qualité d'expert

Expertise. Le CépiDc est sollicité dans des comités nationaux en tant qu'expert, à Santé publique France, la Fédération nationale des observatoires régionaux de santé (conventions). Il est membre de l'Observatoire National du Suicide (DREES), du Comité National d'Experts sur la Mortalité Maternelle. Il est membre du comité des producteurs du centre d'accès sécurisé aux données (CASD). La directrice du CépiDc est membre du comité scientifique du CASD.

Méthodologie – En plus du rattrapage déjà mentionné, les développements méthodologiques, liés en particulier à l'usage de réseaux de neurones dans la classification médicale, menés au CépiDc sont souvent en collaboration, avec : l'ATIH (médecin DIM spécialiste dans le codage médical par apprentissage statistique), le LISEN-Cnrs (ex-LIMSI). Depuis 2024, l'équipe automatisation du CépiDc participe aux échanges animés par le SSP Lab de l'Insee : SSP hub et groupe de travail Insee sur la codification automatique.

Par ailleurs, l'Ined a une historique de travail en collaboration avec Istat sur l'étude des processus causaux de décès issus du codage du système expert Iris. Le CépiDc participe à ce groupe.

Qualité – en tant qu'ONA, le CépiDc est associé à l'animation du réseau qualité de l'Insee.

Etudes/Recherche/Mise à disposition d'extractions – La plupart des exploitations de recherche utilisent la base sur les causes de décès directement via le SNDS. Cependant dans des cas demandant appariement sur trait d'identité, des conventions de mise à disposition spécifiques avec le CépiDc sont mises en place. Elles peuvent accompagner des collaborations en plus du suivi du projet : par exemple sur le suivi des morts maternelles (Enquête nationale confidentielle sur les morts maternelles, Equipe Inserm EPOPé), l'appariement à l'échantillon démographique permanent-Santé (DREES), l'Ined pour divers projets, l'agence de sûreté nucléaire et de radioprotection (suivi de cohortes exposées), registres de pathologies...

3.3 Complémentarité

complémentarité de sa production statistique avec celles d'autres acteurs (SSP ou non).

Complémentarité - La statistique sur les causes de décès est le pendant médical/santé publique de la statistique sur les décès toutes causes et la complémente en premier lieu (même événement générateur, circuit combiné, règlements européens liés, synchronisation...). Elle est issue d'une source administrative, calibrée pour recueillir l'information nécessaire à la construction de la statistique selon le standard international de cette statistique (CIM), encadrée par le code des collectivités territoriales et le code de la santé.

La base de données ainsi construite en alimentant le SNDS complète ce système de données des causes. Pourtant l'appariement réalisé au sein du SNDS pour lier les causes d'un décès à un bénéficiaire souffre

de la non-remontée du statut vital des bénéficiaires dans le référentiel du SNDS pour certaines caisses. Ainsi, alors que pour 98% des décès en France une année (Insee) on est capable d'affecter des causes (chiffres issus de la synchronisation Insee/CépiDc), cela n'est le cas dans le SNDS que pour en moyenne 88%, avec de fortes variations selon les caisses des bénéficiaires. Un projet d'appariement faisant intervenir en amont l'identification au RNIPP via les résultats de l'étape de synchronisation Insee-CépiDc est en cours depuis maintenant plusieurs années, et subit des difficultés pour être effectivement mis en œuvre. Ces difficultés sont liées aux requis des textes et pratiques de mises en œuvre de pseudonomisation du NIR à la Cnam, aux conditions de sécurité des SI propres à chacune des institutions intervenantes, aux rôles de tiers de confiance, aux moyens à affecter à l'opération dans un contexte général contraint, etc..

En revanche, les choses se sont améliorées en ce qui concerne l'enrichissement de l'échantillon démographique permanent EDP-santé tenu par la DREES par les causes de décès. Un avenant à la convention de mise à disposition des données des causes de décès à la DREES au titre de la loi de 1951 a été signé à l'été pour spécifier que le CépiDc met à disposition de la DREES l'extraction des causes de décès des défunts entrant dans l'échantillon démographique permanent, de façon à ce que la DREES puisse enrichir cet échantillon des années récentes. La DREES a procédé à un nouvel appariement entre l'EDP et les causes de décès sur la base des variables les plus « cohérentes » présentes dans les deux bases à savoir les variables issues de la « synchronisation » / alignement avec les décès enregistrés à l'Etat civil. Le taux d'appariement atteint 99% (les rejets proviennent de certains individus qui entrent dans l'EDP via d'autres voies que le RNIPP / état-civil). Ce travail a été mené en septembre 2024. Cet appariement sera prochainement mobilisé par la DREES dans un études et résultats et un document de travail en préparation.

Par construction du SNDS, la Cnil n'autorise pas la mise à disposition du jour de naissance dans les informations sur les décès accompagnant les causes au SNDS. Ceci réduit les possibilités d'appariement à des cohortes ou des enquêtes sur la base des traits identifiants mais non nominatifs disponibles dans la base de données sur les causes de décès, i.e. dates de naissance et de décès, communes de décès, et de naissance. Du fait de la piètre qualité de l'appariement entre le référentiel des bénéficiaires et les causes de décès, il n'y a pas non plus d'intérêt à passer directement par le SNDS. Lorsque ces informations sont nécessaires pour un projet de recherche précis, la mise à disposition des données requiert une autorisation de la Cnil (via avis CESREES) et est systématiquement encadrée par une convention entre l'équipe demandeuse et le CépiDc. Les demandes passent par le guichet unique assuré par la Plateforme de données de santé (HDH). Après autorisation de la CNIL, les projets de recherche demandent une identification RNIPP des observations de leurs projets via le CESP (unité U108/CESP). Cette identification RNIPP est réalisée par l'Insee, qui renvoie les traits d'identification non nominatifs (date et lieu de naissance, décès) permettant ensuite de réaliser l'enrichissement des causes de décès. Tout cela est inscrit dans le code de la recherche [Section 2 : Accès aux données relatives au décès des personnes inscrites au répertoire national d'identification des personnes physiques \(Articles R225-2 à R225-4\)](#)

Exemples de conventions récentes :

-Le projet ORICAMS porté par l'IRSN sur le lien entre exposition professionnelle aux rayonnements ionisants et l'excès de décès par cancer du cerveau ; SPACE sur les professionnels de l'aéronautique ;

-La cohorte E3N-Génération de femmes affiliées à la MGEN et suivies depuis 1990 , + pères et enfants ;

-L'Enquête Nationale Confidentielle sur les Morts Maternelles (ENCMM) menée par l'unité EPOPé de l'Inserm alimente les données mises à disposition du Comité national d'experts sur les morts maternelles. L'enquête s'appuie sur un protocole de repérage large des possibles morts maternelles, sur la base des informations dans les certificats de décès, les bulletins de naissance/ avis de décès enregistrés à l'état

civil, et le retour aux médecins (lesquels sont prévenus lorsqu'ils certifient sous CertDc et déclarent une grossesse liée au décès s'ils acceptent d'être recontactés dans le cadre de l'enquête).

Autres projets démontrant la complémentarité de la source :

-L'enquête Fin de vie, où la base des causes de décès a servi de base de sondage.

-L'Ined porte deux projets : un premier projet sur l'évolution de la mortalité - différentiels par groupe d'âges, de causes, différentiels géographiques et sociaux, pour lequel les données sur les causes de décès mises au SNDS ne sont pas suffisantes car elles ne contiennent pas le jour de naissance nécessaire pour calculer les âges au dernier anniversaire utilisés par les démographes. Le second sur la mortalité des enfants d'immigrés qui s'appuie sur un appariement des bulletins de naissance avec les bulletins de décès et des EAR avec les bulletins de décès, enrichis des causes.

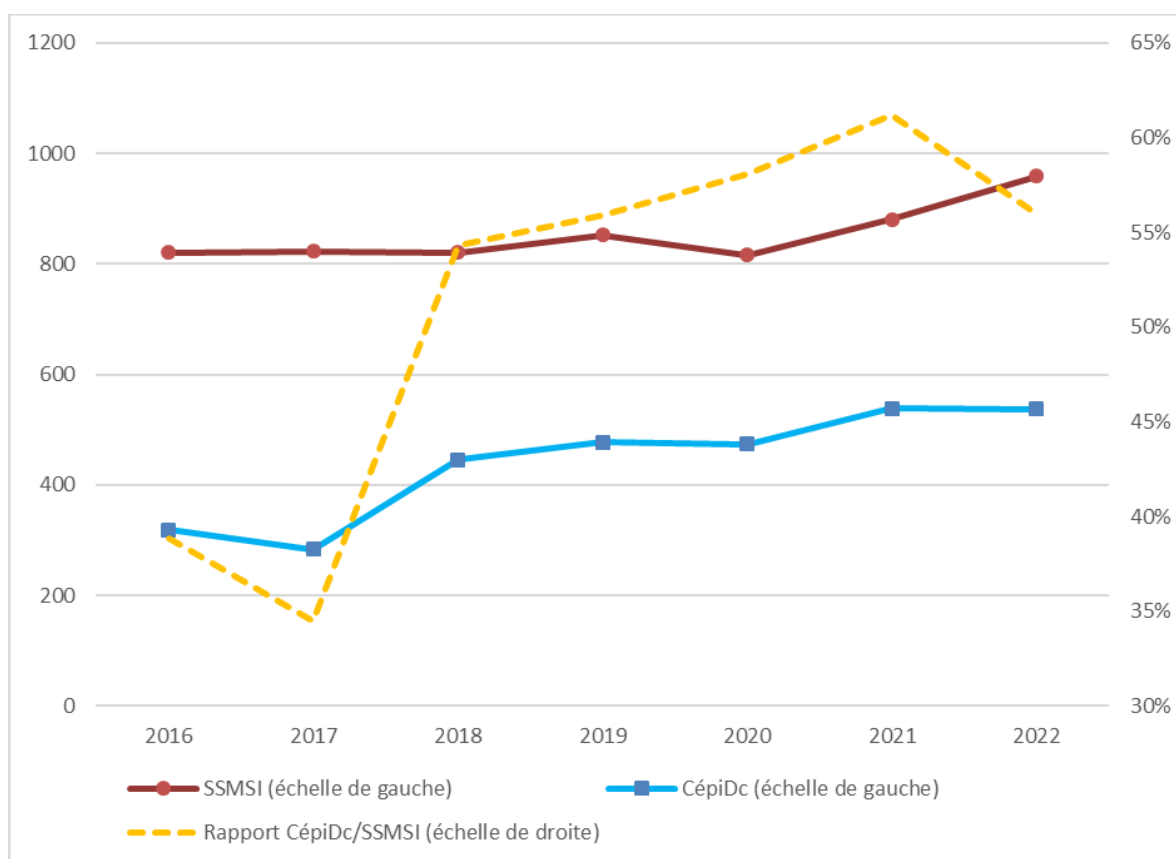
- Les registres en population de morbi-mortalité français suivent tous les cas recensés d'une pathologie donnée. Ils souhaitent accéder aux causes de décès pour un suivi exhaustif et un suivi en routine. La demande a été portée au comité stratégique sur la statistique des causes de décès. Ils entament une demande centralisée d'accès / mise à disposition d'un côté du SNDS et de l'autre du flux d'informations provenant du CépiDc à l'intention de la veille sanitaire (données brutes). Mais la première étape pour les registres est de consolider leurs bases légales de traitement.

Cohérence externe – La statistique des causes de décès en France est comparée lors des publications marronières accompagnant la diffusion de l'année aux statistiques équivalentes produites par les homologues européens (soumis au même règlement) ou internationaux (soumis aux mêmes standards via la CIM): publications des INS, publications académiques ainsi que [Metadata](#) qui renseignent sur les spécificités des pays une année donnée comme pour toute statistique officielle : Metadata pour la France : https://ec.europa.eu/eurostat/cache/metadata/EN/hlth_cdeath_simsd_fr.htm

Au niveau national, d'autres statistiques produites par d'autres acteurs du SSP ou hors SSP, peuvent se rapprocher pour des populations spécifiques de la statistique des causes de décès. Elles diffèrent forcément du fait de définition, de collecte, d'objectifs sensiblement différents mais peuvent être mobilisées pour questionner la cohérence externe des données.

En particulier, le [nombre de victimes d'homicides](#) fiabilisé par le SSMSI dans les logiciels de rédaction des procédures de la police et de la gendarmerie (voir Info rapide n°31) qui recensent les victimes issues des procédures comportant les atteintes volontaires à la vie (homicides intentionnels) et les violences volontaires ayant entraîné la mort sans intention de la donner. Cette définition et le mode de collecte diffèrent de l'information sur la cause telle qu'elle est définie par la CIM – « mort violente avec intention contre la vie d'autrui », déclarée par le médecin au moment du constat de décès selon les circonstances apparentes du décès à ce moment-là, déclaration sous secret médical. On s'attend à ce que l'effectif d'homicides au sens des causes de décès soit inférieur à celui au sens des procédures car s'il y a mort violente mais que le médecin ne sait pas se prononcer sur l'intention ou même l'existence ou non d'une intention et hésite à dire s'il s'agit d'un accident (sans intention) ou d'un homicide, la cause sera un événement dont l'intention est indéterminée. Et les procédures peuvent être requalifiées plus tardivement. En pratique cependant les sens des hétérogénéités régionales et les évolutions temporelles entre les deux concepts peuvent se comparer. Depuis l'année de décès 2018, le CépiDc intègre directement une extraction du SI de l'IML de Paris. A cette occasion, le ratio d'homicides par cause / homicides procédures est passé de 1/3 à plus de 50%. Pour autant les variations annuelles (67% en 2021, 46% en 2022) fortes ne permettent pas de présenter des évolutions temporelles cohérentes avec les chiffres du SSMSI. Des échanges CépiDc SSMSI en cours visent à identifier finement les écarts et

notamment ceux en lien avec une non-remontée d'information au CépiDc lorsqu'une enquête médico-légale est ouverte.

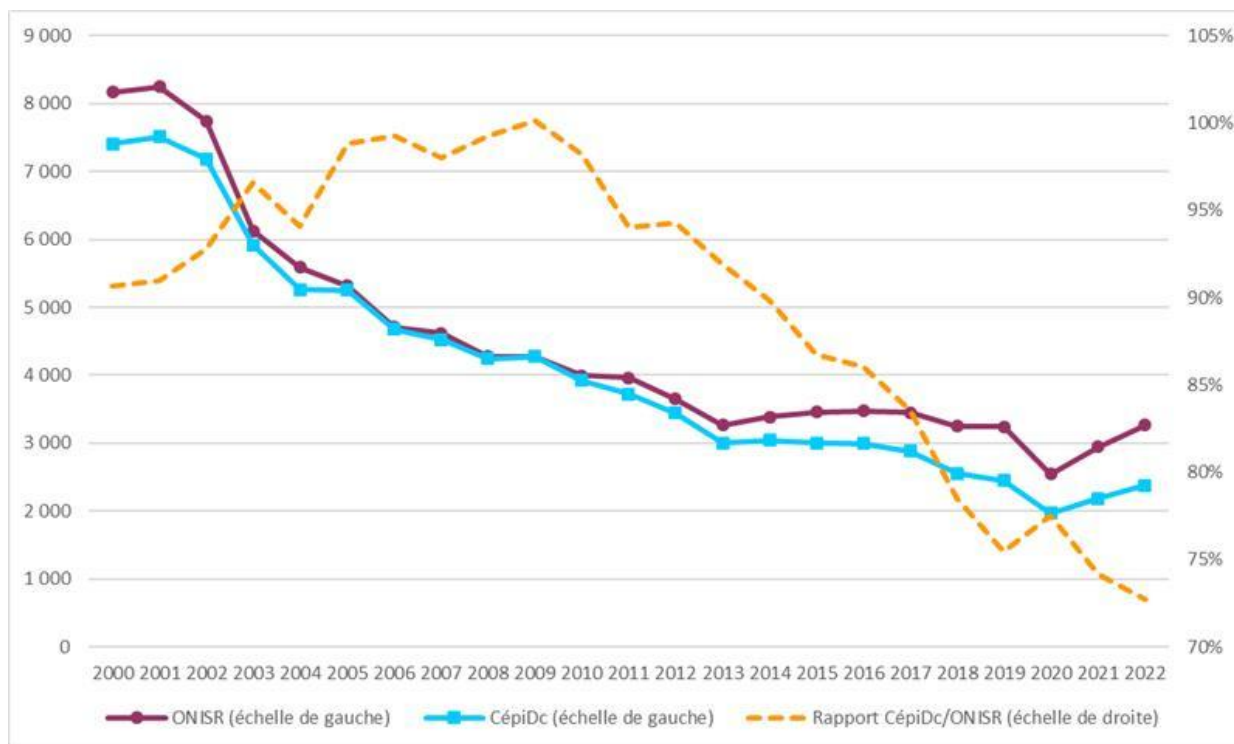


Graphique : comparaison du nombre de victimes d'homicides en France selon les remontées et définitions des causes de décès et selon les statistiques du SSMSI sur les procédures de la police et de la gendarmerie.

Source : SSMSI, CépiDc, calculs CépiDc

Autre source de cohérence externe, les statistiques de l'ONISR sur les [personnes tuées du fait d'un accident de la route](#) (sur le moment ou dans les 30 jours) provenant des déclarations des forces de l'ordre via le Fichier national des accidents corporels. Contrairement aux homicides, où les concepts mesurés diffèrent par nature, la CIM suit les mêmes principes que la statistique sur les accidents du transport sur route, à savoir une différenciation des accidents par le mode de transport de la victime (piéton, vélo, auto...) et celui de ce qui a provoqué l'accident. Les plages des codes de la CIM correspondant le mieux aux accidents de transport sur route correspondent à l'ensemble des accidents ayant lieu sur une voie terrestre ouverte à la circulation publique et impliquant un véhicule (V01-V89) + une partie des séquelles d'accident de transport non spécifié (Y85). Jusqu'en 2011, les séries étaient très proches et évoluaient de façon similaire. A partir de 2012, il y a décrochage. Ceci provient du fait que la règle internationale recommande que si l'implication d'un véhicule n'est pas explicitement spécifiée dans le certificat, celui-ci est codé en « autre accident non spécifié », non comptabilisé en accidents du transport (mécanisme appliqué par le logiciel de codage Iris, batch ou en interactif, et confirmé par le Canada, Italie et le UK). Ce décrochage semble augmenter à l'entrée en vigueur du nouveau modèle de certificat de décès dans lequel les partie 1 et partie 2 sont complétées de cases à cocher (voir après le modèle international et les modèles français), et leurs différences. En pratique, un « accident voie publique » écrit en partie 1 ou 2 conduit à un code d'accident de la route (car c'est ce qui était fait avant la mise en place du modèle international de 2016), mais un terme de traumatisme (différent de accident voie publique) en partie 1

ou 2 + une case cochée « accident » + une case cochée « voie publique » conduit à un code relevant d' « autre accident non spécifié » (X59.9). En effet, l'introduction de la possibilité de cocher des cases semble avoir pour effet que les certificateurs s'abstiennent d'écrire la même information dans l'enchaînement des causes (Partie 1 et 2). Par ailleurs, le modèle français ne reprend pas la question du modèle international « How did the external cause occur ? ». Or c'est la réponse à cette question qui est mobilisée pour caractériser l'accident par les autres pays : « car collision » permettra par exemple d'affecter l'accident à un accident de transports.



Graphique – Comparaison du nombre de décès en France par accidents routiers (selon les causes de décès) et de tués par un accident de la route selon l'ONISR

Sources : ONISR, CépîDc, calculs CépîDc

Enfin, des travaux de recherche ont par le passé comparé les causes initiales de décès aux diagnostics principaux d'hospitalisation tels qu'ils étaient automatiquement définis dans le PMSI, soulignant une similitude modérée entre les concepts – (voir Lamarche-Vadel, A., Pavillon, G., Aouba, A. 2014). Ces travaux n'ont pas été réactualisés.

4 Séries pour lesquelles est demandée la labellisation d'intérêt général et de qualité statistique

(principes n° 11 et 14 du CBPSE).

a) Présentation générale des séries dont la labellisation est demandée.

(Privilégier des intitulés construits de manière cohérente pour toutes les séries (pour mieux faire apparaître les différences et points communs) et facilitant l'appréhension du contenu précis de la série (par exemple, "foyers bénéficiaires du RSA" au lieu de "bénéficiaires du RSA", masses annuelles d'exonérations fiscales" au lieu de "exonérations fiscales").

- **Présentation globale** du jeu de séries proposé et de sa cohérence d'ensemble ; enjeux socio-économiques associés (ou en matière de politiques publiques etc) à l'analyse de ces séries temporelles ; en cas de renouvellement de la procédure de labellisation : présentation et motivation générales des évolutions de

la liste : nouvelles demandes, demandes de passage de la reconnaissance d'intérêt général à la labellisation, séries dont la labellisation n'est pas redemandée.

Comme vu plus haut, la demande de labellisation porte sur la statistique sur les causes de décès, statistique de référence en santé publique, épidémiologie, démographie et statistique depuis la fin du XIX^e siècle. Les causes de décès sont mobilisées dans le cadre de la veille sanitaire, des politiques de santé publique et dans ceux de la statistique et de la recherche en santé. (voir l'article du Courrier des stats décembre 2024 : Coudin, Robert 2024). Collecte (certificats de décès), codification (Classification internationale des maladies) et définition des indicateurs statistiques (cause initiale de décès, causes multiples) suivent les standards internationaux de l'Organisation mondiale de la santé pour assurer la comparabilité dans le temps et dans l'espace des données et des analyses qui en découlent. Cette conformité est obligatoire en Europe, depuis que cette statistique fait partie des statistiques communautaires de la santé publique (voir le manuel 2024 sur les causes de décès d'Eurostat très complet et clair fourni en annexe).

La demande de labellisation porte sur l'indicateur statistique le plus utilisé pour les comparaisons temporelles et internationales, et celui qui est collecté par Eurostat, à savoir la **cause initiale** d'un décès codée dans la classification internationale des maladies.

L'unité statistique est le décès d'une personne survenant en France une année donnée. Les morts-nés ne sont donc pas dans le champ. La CIM et Eurostat précisent la définition d'un décès : « disparition définitive de toute trace de vie à tout moment après une naissance vivante (cessation postnatale des fonctions vitales sans possibilité de réanimation). Cette définition exclut les mortinaissances » (voir les règlements et le manuel sur les causes de décès en annexe).

Les données envoyées à Eurostat concernent l'ensemble des décès occurring sur le territoire de France au sens NUTS FR (c'est-à-dire métropole + DROMs + Saint-Martin), les séries diffusées sur le site du CépiDc et dans les publications accompagnant la diffusion nationale sont restreintes aux décès en France (métropole + DROMs) de résidents en France (au sens Insee de résidents). Il y a en effet chaque année autour de 2000 (sur 640 000) décès de non-résidents en France qui occurrent sur le territoire français (y compris COM sauf Saint-Martin). (2018 : 2202, 2019 : 2277, 2020 : 1719, 2021 : 1755, 2022 : 2153)

La **cause initiale du décès** est un indicateur statistique unique par décès défini dans la classification internationale des maladies. Il correspond à « la maladie ou le traumatisme qui a déclenché la série d'événements morbides conduisant directement à la mort, ou les circonstances de l'accident ou de la violence qui ont produit le traumatisme mortel » (voir le volume 2 de la [CIM](#), [règlement UE 328](#)). Sa détermination se fait grâce à une collecte standardisée du processus morbide (causes of death) et d'autres informations médicales standardisées (voir le modèle de certificat international), ainsi qu'une série de règles de codage spécifiées dans le volume 2 de la CIM, lesquelles sont en grande partie automatisables.

Pour la diffusion, le CépiDc suit les préconisations européennes (Eurostat) et diffuse ces statistiques par regroupement de causes selon une [short-list européenne](#) élaborée sous l'égide d'Eurostat (voir Annexe A du manuel Causes de décès en annexe) et partagée avec une grande partie des autres pays européens.

Liste synthétique des séries¹ (extraite du tableau complet à joindre en complément au dossier (cf. précisions en annexe)). Les séries seront classées : - au premier niveau par thème, - puis le cas échéant par unité statistique, - et enfin par indicateurs portant sur cette même unité statistique. Les indicateurs produits peuvent être des :- effectifs, montants, en valeur ou volume ; - indices d'évolution, en volume, en valeur ; - glissements (d'une année sur l'autre, d'un trimestre donné par rapport au même trimestre de l'année précédente, etc), évolutions moyennes sur une période de référence ; - données brutes ; corrigées de variations saisonnières, corrigées des jours ouvrables, etc.

Il y a deux indicateurs pour lesquels la labellisation est demandée : les **effectifs** de décès par regroupement de causes de la short-list européenne et les **taux standardisés** de mortalité par causes, pour ces mêmes regroupements.

Le taux standardisé de décès selon l'âge correspond à la proportion fictive de personnes décédées dans l'année si la population avait la même structure d'âge qu'une population de référence donnée (concept finalement proche d'une espérance de vie à la naissance). Cette standardisation permet de comparer les taux de décès entre populations n'ayant pas les mêmes structures d'âge soit pour des comparaisons temporelles, en excluant l'effet de l'âge de la population sur la variation des taux, soit pour des comparaisons spatiales entre pays ou régions présentant une répartition par âge différente des habitants. Le taux standardisé selon l'âge et le sexe correspond à la moyenne du taux standardisé selon l'âge des hommes et celui des femmes (population fictive où il y a autant d'hommes que de femmes à tous les âges).

Une population de référence est en pratique une série de pondérations par classe d'âge. Les pondérations mobilisées sur le site du CépiDc en priorité sont celles correspondant à la population standardisée européenne (European Standard Population [Eurostat, 2013]), aussi mobilisées notamment par Eurostat dans les comparaisons internationales. Elles correspondent à la population en 2010 des 27 pays de l'Union européenne, de la Suisse, de l'Islande, de la Norvège et du Liechtenstein (voir la [Fiche méthodologique](#) aussi reportée en annexe).

On présente le taux standardisé par sexe et âge sur la page « grandes tendances <https://www.cepiddc.inserm.fr/donnees-et-publications/grandes-causes-de-deces-en-2022-et-tendances-recentes> » et le taux standardisé seulement par l'âge (standardisation des taux bruts tous sexes) sur l'open-data (<https://www.cepiddc.inserm.fr/donnees-et-publications/interroger-les-donnees-de-mortalite>). Dans une population où les femmes sont plus nombreuses que les hommes aux âges élevés, une catégorie pourtant faiblement représentée sur l'ensemble de la population de référence, la moyenne des taux standardisés masculins et féminins est légèrement supérieure à la standardisation des taux bruts tous sexes. **Nous demandons la labellisation des deux variantes des taux standardisés tous sexes, (différences documentées dans la fiche méthodologique)**, car la standardisation par l'âge uniquement suit les habitudes internationales et la standardisation par l'âge et le sexe est plus adaptée car elle contrôle véritablement des évolutions de structure de la population (espérance de vie plus élevée chez les femmes que les hommes).

b) Critères ayant amené à considérer les séries statistiques comme relevant de l'intérêt général et à les proposer à la labellisation

- ***Modalités d'expression de l'intérêt général*** (Cnis, recueil de besoins, recommandations...) ; existence de demandes pour ces séries pour des usages d'intérêt général (importance dans le débat public, demandes de ministères ou institutions publiques, de corps de contrôles, de journalistes, statistiques de consultation des internautes, chiffres clés repris par l'organisme dans ses rapports...)
- ***Séries (ou données sous-jacentes) contribuant à la production ou aux publications statistiques d'autres organismes*** (statistique publique, Eurostat, OCDE, ONU...), ou répondant à des obligations statistiques légales (fournir les références normatives : loi, règlement, etc.) ; partenariats éventuels dans un processus en coproduction ;
- ***Apport par rapport à d'autres séries*** (de la statistique publique ou non) ; travaux méthodologiques communs autour de ces séries avec d'autres organismes de la statistique publique ou dont les séries sont labellisées/reconnues d'intérêt général par l'ASP.
- ***Autres*** : investissement récent sur la qualité des séries, projet de diffusion ou autre enjeu spécifique...

Le premier critère motivant la demande de labellisation est le fait que la statistique est sous règlement européen, et qu'au titre d'ONA, le CépiDc est tenu de rendre compte de la qualité de la production de cette statistique (cf lettre du DG de l'Insee au PDG de l'Inserm en annexe). Pour ce qui est des taux standardisés, principaux indicateurs contrôlant des évolutions de structure par âge, on note que

n'interviennent dans leur construction que les estimations de population annuelles produites par l'Insee et répondant au règlement statistique démographique.

Au-delà des partenaires et des catégories déjà mentionnés plus haut, les utilisateurs principaux de ces séries sont les médias grand public (les marronniers Etudes et Résultats et Bulletin épidémiologique hebdomadaire sont largement repris par la presse à leur sortie), les acteurs en santé publique nationaux et régionaux (SpF, ARS, ministère de la santé, ORS), les épidémiologistes, démographes, chercheurs en santé et utilisateurs du SNDS, statisticiens qui vont trouver ici les évolutions de référence (Ined, Insee, DREES,...). Puis, au niveau international via l'OMS la communauté de recherche et de politique en santé publique, et via Eurostat, la DG santé, le *Joint Research Center* pour des projets de recherche notamment *European Cancer Information System*, politiques, administrations nationales, universités, recherche... (voir Manuel sur les causes de décès 2024).

Enfin, il est clair que les travaux de rattrapage et de refonte du processus de production de la statistique récents motivent aussi la demande de labellisation afin de les mettre en évidence et d'assurer leur visibilité et plus généralement la visibilité de ces séries dans le monde de la recherche et de la statistique en santé puisque les données sont dorénavant assez fraîches pour être véritablement informatives.

Pour l'instant, il n'y a pas d'exposition de l'intérêt général devant le Cnis. Cela provient certainement de l'historique de la production de la source, hors SSP dès lors qu'elle a été confiée à l'Inserm. Pour autant, on peut avancer plusieurs arguments en faveur de la mise en place d'une exposition de ce type : le CépiDc est un « petit » producteur statistique qui pourrait bénéficier du canal Cnis comme lieu d'échange avec les utilisateurs, et aussi les statisticiens, d'autant plus que le modèle international de certification médicale tel que défini dans la CIM est désormais un petit « questionnaire » statistique, avec des problématiques de statistique d'enquête (on y reviendra).

c) Description et justification des principaux concepts, unités, nomenclatures, champs, utilisés, cohérence interne et externe (avec le SSP ou d'autres références)

Les descriptions doivent être compréhensibles par un utilisateur non spécialiste du champ, et lui permettre de mettre les séries en regard d'autres données. Elles doivent tenir compte du fait que les séries ont une dimension temporelle. Elles peuvent renvoyer pour plus de détail à de la documentation en ligne. Les notions utilisées doivent avoir un sens pour l'intérêt général (et non refléter uniquement des modalités de gestion). Selon les situations, la présentation peut être globale ou factorisée par groupes de séries.

- **Unités statistiques, nomenclatures et critères de ventilation, concepts** (statistiques ou du domaine) :
 - définitions, évolutions et rétropositions éventuelles ;
 - cohérence des définitions (ou explications et justification des écarts) : entre séries, avec des définitions utilisées dans d'autres sources ou publications statistiques (de la statistique publique ou non), avec d'autres références (juridiques...)...
- **Géographie** :
 - localisation géographique prise en compte (résidence, travail, siège de l'entreprise...)
 - description du champ pour les différentes unités statistiques (y compris traitement des DOM, COM, prise en compte de résidents ou établissements à l'étranger...)
 - gestion des évolutions de géographie (y c. rétroposition).
 - explicitation des choix géographiques (au regard des compétences de l'organisme, des enjeux socio-économiques associés aux séries, des politiques publiques associées à différents niveaux géographiques...).
- **Temporalité de la série et motivations associées** :
 - périodicité de production (annuelle, trimestrielle, mensuelle...)
 - référence temporelle (sur une période, en fin de période, etc.) de la série avec ses dates de début et de fin ;
 - délai de disponibilité (pour les résultats provisoires comme pour les résultats définitifs) ;
 - date d'extraction ou de mise à jour des données ;
- **Le cas échéant : critères de ventilation des séries et motivations associées**

Tous les concepts (nomenclatures y compris définition d'indicateur statistique de cause initiale, définitions de la géographie, ventilation par âge, sexe, ...) proviennent du règlement EC 1338/2008 et de son règlement d'application EU 328/2011, lesquels réfèrent directement aux concepts définis dans la

Classification internationale des maladies de l'OMS. La labellisation est demandée pour les séries depuis 1979, année à partir de laquelle le CépiDc a publié les séries selon la short-list européenne.

Les séries sont ventilées par département du domicile du défunt. Elles sont disponibles pour la France métropolitaine depuis 1979, à laquelle s'ajoutent la Martinique, la Guadeloupe et la Réunion depuis 2000, la Guyane depuis 2001 et Mayotte depuis 2014. A l'échelle de diffusion géographique envisagée, la seule révision géographique concerne le redécoupage des régions en 2016. Les indicateurs sont diffusés dans les deux nomenclatures sur l'ensemble de la période.

Les séries d'effectifs et de taux standardisés annuels sont produits une fois par an. Le règlement européen requiert la livraison des données dans les 24 mois suivant l'année de référence. Un *gentlemen agreement* (largement suivi par les pays européens) propose une livraison des données dans les 18 mois après l'année de référence. En France, les données 2022 ont été envoyées à Eurostat fin septembre 2024, puis diffusées au niveau national le 8 octobre. Le calendrier de diffusion pour 2023 est fin juin 2025. La France rejoindra donc le *gentlemen agreement*.

Les effectifs de décès sont définitifs à la date de leur publication. Les taux standardisés sont provisoires tant que les estimations de population de l'Insee sont révisées : ils deviennent définitifs 48 mois après l'année de référence considérée.

Dans les publications marronnières ([Etudes et résultats 2021](#), [2022](#)), des premières estimations de tendances sur des agrégations plus larges pour l'année N+1 sont fournies et commentées, ainsi qu'un retour sur l'écart entre les chiffres de l'année N et les estimations produites l'année d'avant.

5 Etat des lieux de la mise en œuvre des recommandations émises à la précédente labellisation

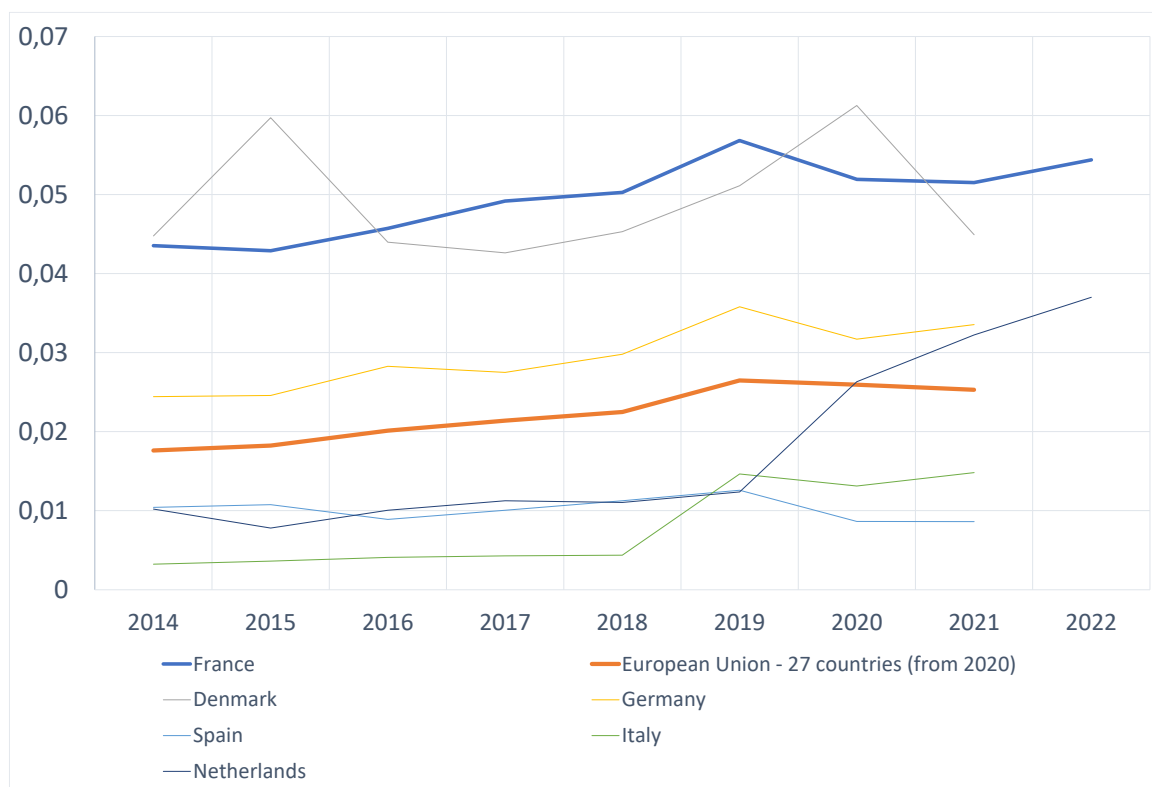
état des lieux de la mise en œuvre des recommandations émises lors de la labellisation précédente (se rapporte au principe n° 4 du CBPSE)

Recommandation 1 : Indépendance professionnelle - il importe que cette situation d'indépendance professionnelle perdue. Pour être en pleine conformité avec le règlement statistique 223/2009 révisé, il conviendrait à l'avenir de clarifier les procédures de recrutement et de nomination du directeur du CépiDc, comme cela était demandé dans le courrier du Directeur général de l'Insee du 6 juillet 2016.

Le CépiDc a changé de directeur/trice le 1^{er} septembre 2022. Le PDG de l'Inserm a témoigné de la procédure de recrutement appliquée à cette occasion en application des recommandations de l'article 5bis point 4 du règlement CE 223/2009 dans une lettre adressée à la présidente de l'ASP en septembre 2022 : publicité de l'ouverture du poste, profil professionnel, rencontres, avis de l'Insee, conditions d'emploi (lettre jointe en annexe + fiche de poste 2022). Elise Coudin, actuelle directrice, est en détachement de l'Insee sur contrat à l'Inserm jusqu'au 31/08/2025. Le poste de directeur/trice du CépiDc est ainsi actuellement ouvert. La fiche de poste a été publiée en janvier sur Choisir l'emploi public et dans la campagne interne de l'Inserm. L'ouverture de ce poste à l'Inserm a aussi été annoncé en comité de direction de l'Insee et la fiche de poste est accessible sur l'intranet de l'Insee. Cette fiche de poste indique les compétences statistiques et le niveau de compétence requis pour occuper le poste. Elle indique aussi les modalités de candidatures et les modalités de sélection : « Modalités de sélection : un comité de sélection sera chargé d'examiner les dossiers et de procéder aux auditions des candidats retenus à l'oral. Le CépiDc étant soumis au Règlement européen 223 (règlement statistique), le comité de sélection inclura un représentant du service statistique public. » (voir Annexe fiche de poste 2025)

Enfin, le fait que ce recrutement doive suivre le code des bonnes pratiques en matière de statistique européenne est inscrit dans le règlement intérieur de l'unité CépiDc (Annexe).

Recommandation 2 : Le Comité encourage le CépiDc à contenir, voire à faire diminuer le nombre d'observations de décès ayant une cause inconnue, au besoin par une phase l'imputation statistique et en veillant à impliquer les acteurs déficitaires de la collecte.



Graphique – Part des causes inconnues ou mal définies : autre mort subite, mort inattendue et causes inconnues, R96-R99, dans les décès (%)

Source – eurodatabase, extraction décembre 2024

Sur ce point, les choses ne se sont pas sensiblement améliorées, la France reste un des pays en Europe où la part de causes considérées comme inconnues est la plus élevée, dépassant 5%. Parmi les raisons, - un taux de collecte autour de 98% conduisant à imputer 2% des décès, et ces cas, en suivant la recommandation de la CIM, sont inclus avec le code R99 (cause inconnue ou non précisée). A noter que la précédente proposition du Label de redresser les valeurs manquantes sur la base des valeurs observées ne suit pas la recommandation de la CIM (même si un redressement pas sexe et âge serait certainement suffisant pour identifier des morts violentes). Ce travail peut en revanche être mené dans le cadre d'études ou d'analyses spécifiques. Des travaux proches sont notamment menés dans le cadre du « Global Burden of diseases » (travaux associant Santé publique France pour la France) où les causes mal définies (au sens plus large) sont réaffectées à des catégories ou chapitres plus interprétables.

- Une recommandation de l'OMS à partir de 2016, en pratique dans les pays à partir de 2019 et de ne pas retenir les codes peu informatifs I46.0 (arrêt cardiaque réanimé) et I46.9 (arrêt cardiaque sans précision) en cause initiale. Ces cas passent en R99 à partir de 2019, cela concerne environ 4700 cas. Ce changement (que l'on voit en moyenne sur l'Europe) explique la hausse de la proportion entre 2018 et 2019, à « définition » d'avant 2016, on resterait en dessous des 5%. Pour la France, l'effet de cette rupture est chiffrée dans le rapport de production 2018/2019 <https://www.cephidc.inserm.fr/documentation/rapport-de-production-annees-de-deces-2018-et-2019-donnees-definitives-document-de-travail-du-cepidc-n32023>.

La précédente labellisation s'est déroulée juste avant la mise en œuvre d'un nouveau modèle de certificat, lequel se conforme à la CIM dans sa version 2016 en demandant de collecter en plus des parties sur les

causes médicales, des informations médicales complémentaires pour mieux catégoriser les causes. Ce nouveau certificat s'est diffusé à partir de début 2018. A cette occasion, a été introduit un volet médical complémentaire (VMC). Celui-ci doit être rédigé en cas d'investigations médicales ou médico-légales post-mortem par le médecin légiste qui les exécute. Il suit le même modèle de certificat et se complète sur la plateforme de certification électronique. Les VMC sont collectés à partir du 1er janvier 2018. Les Instituts médico-légal (IML) ne les remplissent en pratique que très rarement. Hors IML de Paris (cas particulier sur lequel on revient après), le CépiDc a reçu 45 VMC en 2018, 143 en 2019, 731 en 2021, et 598 en 2022, soit très peu. C'est un échec. L'alternative élaborée pour l'IML de Paris s'est avérée plus efficace. Dans ce cas, l'IML envoie au CépiDc un extrait de son système d'information. Ces données secondaires ne correspondent pas au type de recueil et au format requis pour coder en bonne et due forme dans la CIM (ce qui pose question à l'équipe de codage) mais elles constituent des sortes de proxy, en particulier utiles pour capter des cas d'homicides. Les bases finales contiennent 1800 certificats issus de ce système pour 2018 et autour de 2500 par an depuis 2019.

Enfin, un point soulevé par la mission des inspections générales sur les statistiques de décès concerne le manque de compréhension par le certificateur de ce qui est attendu de lui pour décrire le processus morbide, du fait de manque de formation, de documentation, d'une impression à tort qu'il s'agit d'une démarche uniquement médico-légale. Ceci reste vrai malgré des efforts de documentation sur la plateforme de certification électronique, un article à l'attention des médecins dans Presse formation médicale (voir Coudin, Fouillet 2025 dans la Presse formation médicale en annexe), des échanges avec la DGS (dans le cadre de l'extension de la certification aux infirmiers), voire l'ordre des médecins. En outre, le nouveau modèle international de certificat complète la description attendue du processus morbide par des questions spécifiques (voir après les modèles). Il s'avère que la traduction de ces questions en français, ainsi la façon dont elles s'articulent avec d'autres questions posées au niveau national interrogent la bonne compréhension de ce qui est attendu du certificateur. Par exemple, « *manner of death* » sensé capter l'intention derrière une mort violente (intention de nuire à autrui, intention dirigée envers soi-même, absence d'intention, cas de doute avec intention indéterminée) est traduite par « circonstances apparentes de décès » et confond la case relative à une « intention indéterminée ou ne pouvant pas être déterminée » et celle « inconnue ». La spécification sur la raison de la mort violente (*how did the external cause occur* ?) n'est pas présente dans le certificat français. La question sur le lieu institutionnel du décès qui n'est pas requise par le questionnaire international et qui est certes judicieuse, intervient au milieu de l'interrogation statistique nécessaire au codage en CIM. Or l'ordre dans lequel les questions sont posées peut avoir un impact sur les réponses. En particulier sur les causes externes, on observe et documente les effets en pratique du changement de modèle de certificat (plus d'intentions indéterminées car confusion avec circonstances de décès inconnues, moins de texte décrit dans les causes si une case dit la même chose, par exemple « suicide »). Deux suggestions sur lesquelles le CépiDc demande discussion et avis du comité du Label. A l'instar des BEC qui sont présentés (pour la partie statistique) au Cnis, serait-il utile d'envisager d'ajouter dans les volets médicaux des certificats de décès une cartouche faisant mention de l'intérêt général de l'enquête statistique et renvoyant à la labellisation le cas échéant ? Par ailleurs, pourrait-il y avoir un endroit où statisticiens, acteurs de la santé, utilisateurs, soient représentés et puissent « conseiller/valider » le recueil de l'information statistique du volet médical comme un questionnaire ? Le Cnis serait-il l'endroit ?

Recommandation 3 : S'agissant du projet de dématérialisation pour la certification électronique, le Comité encourage le CépiDc à développer toutes les actions favorisant le mode de collecte dématérialisé sachant qu'il s'agit d'un projet concernant de très nombreux acteurs. Il est essentiel qu'il s'agisse vraiment d'une dématérialisation totale pour le médecin déclarant, de façon à ne pas alourdir la charge de ce dernier médecin et à supprimer le frein constaté actuellement pour la certification en ligne. Par ailleurs, le Comité encourage vivement la conduite de recherches pour aboutir à une plus grande efficacité de la codification automatique sur les certificats électroniques (taux de codage automatique actuel de 3 % particulièrement faible).

Depuis 2017, la certification électronique a largement progressé : fin 2024, la moitié des certificats de décès étaient dématérialisés, contre 27% à l'été 2020 (rapport mission interIG). Ceci est dû à plusieurs facteurs, une accélération de la dématérialisation après 2020, un décret instituant l'obligation de certification par voie électronique dans les établissements de santé, et en général sauf exception ([modification des articles R2213-1-2 et R2213-1-4 du code des collectivités territoriales](#)), le raccordement des établissements de santé à la plateforme de certification électronique ([CertDc](#)) et des mairies au [hubEE](#), leur permettant de recevoir par voie électronique les volets administratifs des certificats de décès remplis sur CertDc (sinon le médecin devait imprimer le volet administratif), le travail de la DGS sur le terrain, auprès des ARS notamment, pour promouvoir la certification électronique, la mise en place d'une plateforme de certification pour les mobiles...

Pour ce qui est du codage automatique par le système de règles Iris, la mission interIG dressait fin 2020 le constat suivant « seuls 52% des certificats de 2020, exploités au 10 décembre, avaient pu être codés de façon automatique, (...) 58% pour les certificats papiers et 36% pour les certificats électroniques. » Depuis, un travail simplification des règles de codage (affectation de la cause initiale la plus probable, laquelle ne sera revue manuellement que dans des cas spécifiques, dite « approche des choix de codes »), un renvoi vers le prestataire de numérisation/ saisie des certificats électroniques non codés automatiquement, de façon à ce qu'ils puissent subir le même type de standardisation que les certificats papier et un travail sur le dictionnaire de termes, conduisent à monter le taux de codification automatique par batch du système de règles Iris à 61-62% sur l'ensemble des certificats d'une année, ie 62% sur les certificats papiers et 61% sur les certificats électroniques (après standardisation le cas échéant). Si l'on ajoute l'autre approche d'automatisation par prédiction d'algorithmes de deep learning introduite en production régulière à partir de 2021 on arrive à un taux de codage automatique d'environ 87%. Ces éléments sont suivis et présentés dans les rapports de production annuels des données ([2018-2019](#), [2021](#), [2022](#)). Le rapport de production des causes de décès en 2022 est fourni à titre d'exemple en annexe.

Recommandation 4 : Il importe de maintenir le potentiel productif du service en recrutant des agents de profil idoine et d'accélérer le traitement automatique de l'information pour assurer le respect du calendrier de diffusion à Eurostat à l'échéance obligatoire de 24 mois.

Après une période de manques de moyens au plus fort en 2020, grâce aux financements supplémentaires ayant accompagné le plan de rattrapage et de refonte de la statistique sur les causes de décès ainsi qu'à la remise au niveau de 15 postes d'ITA par l'Inserm sur son plafond actuel d'emplois, il a été possible de consolider l'équipe de codage (expertise dans la classification, connaissances médicales), de constituer une équipe automatisation/datascience (mise en place, maintenance des algorithmes prédictifs de deep learning), d'avoir les compétences pour piloter statistiquement la production dans le respect des délais (mise en place d'une stratégie de codification, statisticien dans le pôle production, orientation « validation statistique » du pôle diffusion, détachements de cadres de la statistique publique au CépiDc). Le CépiDc respecte le calendrier de diffusion à Eurostat depuis fin 2023, avec en 2022 des données 2020 diffusées en temps et en heure, des données provisoires 2018 et 2019 puis définitives diffusées en septembre 2023, des données 2021 diffusées en décembre 2023.

Recommandation 5 : À terme et en tirant profit des évolutions du dispositif, le Comité encourage le CépiDc à améliorer le délai de publication des résultats, Le Comité suggère la cible de 12 mois après l'année de référence, qui est un standard courant pour des données annuelles et correspond d'ailleurs au délai de transmission des données Insee sur la population par âge, sexe et région.

Les données 2022 ont été diffusées à Eurostat en septembre 2024, soit à T+21 mois. Il est prévu de diffuser les données 2023 à T+18 mois. La France rejoindrait ainsi le calendrier de la plupart des pays européens, lesquels diffusent à T+18 mois dans le cadre d'un *gentlement agreement* avec Eurostat (21 pays sur 32 pays pour 2022). La cible de 12 mois après l'année de décès n'est pas atteignable actuellement, ce point

a été abordé lors de la réunion du comité stratégique de février 2025 (voir CR en annexe dont on reporte ici les conclusions) :

« Pour les campagnes suivantes, il s'avèrera cependant difficile de soutenir le rythme trimestriel de réduction du délai de diffusion. La campagne de production d'une année doit en effet s'articuler avec la fin de collecte et avec le calendrier de synchronisation annuelle avec les décès de l'Insee. Pour 2026 (données 2024), on pourrait gagner encore un mois ou deux mais une livraison des données définitives à terme avant T+15 mois serait trop optimiste sans une profonde révision de l'étape de synchronisation. En revanche, des données provisoires non synchronisées pourraient, elles, être disponibles plus rapidement. La question étant d'évaluer sous quel format et à quels besoins elles pourraient répondre. Des estimations globales mettant en évidence les grandes tendances de l'année N+1 sont déjà disponibles dans l'Etudes et résultats qui accompagne la sortie des données de l'année N. Ces estimations pour l'année N+1 sont aussi possibles si la publication de l'année N se fait entre N+18 mois et N+15 mois.

Le comité stratégique acte que le calendrier de diffusion des causes de décès données finalisées ne peut pas être réduit à moins de T+16-17 mois dans les prochaines années. Ce point pourra être révisé lorsque la couverture de la certification électronique sera bien plus importante et/ou dans le cas d'une profonde refonte de l'étape de synchronisation. Le comité stratégique note que des estimations provisoires des causes de décès de l'année N pourront en revanche être disponibles à T+6 mois, sur la base d'estimations pour corriger des biais induits par les trous de collecte et avant synchronisation avec les décès Insee, comme cela s'est déjà fait à T+12 mois puis T+9 mois les deux dernières années. Il semble qu'il n'y a pas d'intérêt spécifique à alimenter le SNDS de ces estimations. Mais si les besoins sont clairement identifiés pour un projet de recherche particulier, les données conduisant à ces estimations pourraient être mises à disposition, selon un cadre juridique à expertiser. Ces données devraient alors être utilisées avec des pondérations.

Le CépiDc indique effectuer cette année un travail d'articulation des relances vers les ARS pour accélérer la réception des certificats papiers. La DGS propose son aide pour sensibiliser les ARS, mais aussi le ministère de l'Intérieur pour atteindre les mairies via les Préfectures. »

Recommandation 6 : Il faudra signaler clairement, notamment par un affichage sur le site du CépiDc, que la labellisation porte exclusivement sur les statistiques d'effectifs de décès. En revanche, les autres indicateurs (taux standardisés) qui résultent du rapprochement avec d'autres sources et reposent sur des définitions et conventions spécifiques, sont hors du champ de la labellisation des données.

La mention avait été reportée sur le site internet du CépiDc, puis enlevée lors du non-renouvellement de la labellisation. La demande actuelle est de labelliser des séries d'effectifs et de taux. En fonction de l'issue de cette demande, on pourra préciser l'état de la labellisation, sur le site internet en faisant apparaître un logo de statistique publique si c'est possible, comme ce que fait par exemple l'ONISR.

Recommandation 7 : Il conviendrait d'afficher un calendrier de diffusion des résultats sur le site. De 24 mois dans un premier temps, ce délai affiché sera raccourci au fur et à mesure des améliorations de la production des résultats. Il est attendu un courrier d'engagement de publication signé du Président-directeur général de l'Inserm et du Directeur du CépiDc.

Le calendrier de diffusion est disponible [ici](#).

Recommandation 8 : Il conviendrait d'améliorer la documentation du site sur les points suivants dans le souci de rendre les données mieux accessibles aux utilisateurs :

La documentation a été améliorée, avec un [site internet](#) revu et complété. Une page de synthèse sur les [grandes causes de décès](#) (et les principales agrégations) complète l'outil [open-data](#). Elle fait référence aux dernières statistiques diffusées ainsi qu'aux publications qui les accompagnent. Elle fait aussi référence à la méthodologie suivie (rapports de productions, définition des indicateurs), précise les champs, sources et la nomenclature mobilisée.

- **Notice pour la compréhension du champ couvert ;**

La [fiche méthodologique](#) sur les indicateurs publiée sur le site du CépiDc précise aussi les champs couverts par chaque diffusion (présente en annexe), voir aussi la [page introductive](#) à l'outil d'open-data.

- **Nomenclatures retenues (avec différents niveaux d'arborescence possibles et en soulignant l'intérêt des niveaux agrégés pour des séries plus robustes en évolution sur les causes médicales de décès) ; explication des découpages géographiques (préciser que les grandes villes sont de fait les communes et non les agglomérations, par exemple) ;**

Le CépiDc diffuse désormais les données systématiquement en suivant les [agrégations de causes de la shortlist européenne](#) (voir aussi <https://www.cephidc.inserm.fr/note-dinformation-0>), il s'agit donc des mêmes agrégations que celles que l'on retrouve sur le site d'Eurostat. Ces éléments et les références sont indiqués sur le site internet. Les causes de décès sont codées dans la CIM 10. Le site internet introduit la [classification internationale des maladies et ses versions](#). Les découpages géographiques sont tout-à-fait standards, car définis dans le règlement européen (NUTS FR).

- **Dictionnaire des variables et documentation méthodologique qui gagnerait à reprendre les définitions du règlement européen et devrait présenter le principe des algorithmes de codage, notamment pour la détermination de la cause initiale de décès ;**

Le site internet intègre des pages décrivant en grandes lignes les étapes de la production des causes de décès ([finalités/cadre](#), [collecte](#), [traitements](#)). Il fait notamment lien au document les « [Statistique sur les causes de décès de A à Z](#) » qui résulte d'un travail de documentation entre le CépiDc et l'unité qualité de l'Insee mené en 2022, et à un article dans le [courrier de statistiques de décembre 2024](#), qui présente l'historique, la spécificité de la source et de ses traitements. Par ailleurs, le CépiDc a lancé une collection de documents de travail, mis à disposition sur son site internet. Et depuis les données 2018/2019, le CépiDc produit pour chaque campagne de production un rapport de production qui documente les spécificités de la production annuelle, la méthodologie retenue, ses évolutions, et des éléments d'évaluation. Ce rapport contient le dictionnaire des variables des tables mises au SNDS : [2018/2019](#) ; [2021](#) ; [2022](#) (dernière édition en annexe). A noter aussi la [fiche méthodologique](#) sur les principaux indicateurs (champs, définitions, méthodologie).

- **Citation explicite de l'envoi à Eurostat dans le cadre du règlement 328/2011 et de la possibilité de comparaisons internationales avec un lien vers la partie concernée du site d'Eurostat, voire du site de l'OMS ;**

Fait. Voir <https://www.cephidc.inserm.fr/le-cephidc/missions>

Accès aux données européennes : <https://www.cephidc.inserm.fr/donnees-et-publications/grandes-causes-de-deces-en-2022-et-tendances-recentes>

<https://www.cephidc.inserm.fr/donnees-et-publications/interroger-les-donnees-de-mortalite>

- **Alertes sur les ruptures de séries et renvoi aux éventuelles études d'impact ;**

Une alerte générale sur le risque de rupture de série est indiquée ici <https://www.cephidc.inserm.fr/note-dinformation-0>. Ce point pourra être amélioré dans le futur, un travail d'état des lieux des principales ruptures de séries et de leurs raisons est en cours.

- **Meilleure explicitation de la façon dont est assuré le secret statistique.**

Idem, il y a une mention générale est faite sur le fait que les données de l'open-data sont en partie floutées pour assurer le secret, mais la méthodologie choisie conduit à des biais systématiques sur les décès

sensibles comme les suicides et va être changée, ce qui fournira l'occasion de la documenter plus précisément (voir partie dédiée).

6 Description synthétique du processus de production « amont » des données administratives

acte générateur (texte), processus d'enregistrement, codification de la donnée et traitements de gestion. Description du rôle des différents acteurs (fournir des schémas du processus le cas échéant). Présence de normes, procédures, consignes, contrôles ou autres éléments concourant à la qualité des données amont. (se rapporte aux principes n° 8 et 12 du CBPSE)

6.1 Acte générateur et modèles de certificats

Les statistiques sur les causes de décès sont le pendant des statistiques de décès toutes causes. Elles proviennent du même événement générateur à savoir un médecin constatant un décès. Pourtant suivant les finalités des usages et les modes de transmission de l'information (papier /électronique), les informations collectées/communiquées à chaque organisme ne sont pas toujours les mêmes et les circuits de remontée de l'information diffèrent en partie. Ceci conduit à des délais de transmission de l'information différents et à des traitements de dédoublonnages et synchronisation spécifiques à mettre en œuvre. Aujourd'hui, chaque décès ayant lieu sur le territoire donne lieu à la rédaction d'un certificat de décès par un médecin². La validation de ce certificat est nécessaire pour fermer le cercueil et procéder à l'inhumation. La forme de ce certificat et les finalités d'usage des informations qu'il contient sont définies dans le Code général des collectivités territoriales (Article [L2223-42](#) et Articles [R2213-1-1](#)). Un volet administratif avec les noms et prénoms est destiné aux opérateurs funéraires et à la mairie qui dressera l'acte de décès, et un volet médical confidentiel sans les noms et prénoms du défunt est transmis à l'Inserm. Les finalités d'usage du volet médical sont uniquement statistiques et sanitaires.

Le modèle de certificat est établi par le ministère chargé de la santé ([arrêté du 29 mai 2024](#)), il suit les recommandations/ standards internationaux décrits dans la classification internationale des maladies et repris par les règlements européens (mais s'en écarte aujourd'hui en pratique par certains aspects mentionnés ci-dessus). Il différencie les décès des personnes âgées d'au moins un an (certificat général), et ceux des personnes, nées vivantes (à plus de 22 semaines d'aménorrhée ou pesant plus de 500 g), décédées avant le 365^e jour (certificat infantile).

On reporte ci-dessous les volets médicaux des deux types de certificats en vigueur aujourd'hui, ainsi que le modèle international issu de la CIM (les certificats de décès complets avec le volet administratif sont présents en annexe).

² Une expérimentation est en cours pour autoriser la certification aux infirmiers dans des cas précis suite au [décret n° 2024-375](#). Elle sera généralisée / prolongée à compter du 26 avril 2025.

7.1.1 International form of medical certificate of cause of death

Additional data, that might be necessary for the reporting system of countries, can be added to the certificate. It should not replace the information shown below.

<i>Administrative Data</i> (can be further specified by country)																			
Sex	<input type="checkbox"/> Female			<input type="checkbox"/> Male			<input type="checkbox"/> Unknown												
Date of birth	D	D	M	M	Y	Y	Y	Y	Y	Date of death	D	D	M	M	Y	Y	Y	Y	Y
Frame 1: Medical data: Part I and II																			
1 Report disease or condition directly leading to death on line a Report chain of events in due to order (if applicable) State the underlying cause on the lowest used line		Cause of death						Time interval from onset to death											
		a																	
		b	Due to:																
		c	Due to:																
d	Due to:																		
2 Other significant conditions contributing to death (time intervals can be included in brackets after the condition)																			
Frame 2: Other medical data																			
Was surgery performed within the last 4 weeks?						<input type="checkbox"/> Yes		<input type="checkbox"/> No		<input type="checkbox"/> Unknown									
If yes please specify date of surgery						D	D	M	M	Y	Y	Y	Y						
If yes please specify reason for surgery (disease or condition)																			
Was an autopsy requested?						<input type="checkbox"/> Yes		<input type="checkbox"/> No		<input type="checkbox"/> Unknown									
If yes were the findings used in the certification?						<input type="checkbox"/> Yes		<input type="checkbox"/> No		<input type="checkbox"/> Unknown									
Manner of death:																			
<input type="checkbox"/> Disease			<input type="checkbox"/> Assault			<input type="checkbox"/> Could not be determined													
<input type="checkbox"/> Accident			<input type="checkbox"/> Legal intervention			<input type="checkbox"/> Pending investigation													
<input type="checkbox"/> Intentional self harm			<input type="checkbox"/> War			<input type="checkbox"/> Unknown													
If External cause or poisoning:						Date of injury		D	D	M	M	Y	Y	Y	Y				
Please describe how external cause occurred (If poisoning please specify poisoning agent)																			
Place of occurrence of the external cause:																			
<input type="checkbox"/> At home		<input type="checkbox"/> Residential institution		<input type="checkbox"/> School, other institution, public administrative area			<input type="checkbox"/> Sports and athletics area												
<input type="checkbox"/> Street and highway		<input type="checkbox"/> Trade and service area		<input type="checkbox"/> Industrial and construction area			<input type="checkbox"/> Farm												
<input type="checkbox"/> Other place (please specify):						<input type="checkbox"/> Unknown													
Fetal or infant Death																			
Multiple pregnancy						<input type="checkbox"/> Yes		<input type="checkbox"/> No		<input type="checkbox"/> Unknown									
Stillborn?						<input type="checkbox"/> Yes		<input type="checkbox"/> No		<input type="checkbox"/> Unknown									
If death within 24h specify number of hours survived						Birth weight (in grams)													
Number of completed weeks of pregnancy						Age of mother (years)													
If death was perinatal, please state conditions of mother that affected the fetus and newborn																			
For women, was the deceased pregnant?						<input type="checkbox"/> Yes		<input type="checkbox"/> No		<input type="checkbox"/> Unknown									
<input type="checkbox"/> At time of death						<input type="checkbox"/> Within 42 days before the death													
<input type="checkbox"/> Between 43 days up to 1 year before death						<input type="checkbox"/> Unknown													
Did the pregnancy contribute to the death?						<input type="checkbox"/> Yes		<input type="checkbox"/> No		<input type="checkbox"/> Unknown									

Le volet médical dans sa dérivation française suit le principe suivant. Dans une première partie relative aux **causes (Partie 1)**, le médecin écrit sur quatre lignes l'enchaînement causal des maladies ou conditions qui ont conduit au décès (aussi appelé « processus morbide »), de la cause immédiate à inscrire en première ligne, à la cause initiale, celle qui a initié le processus conduisant au décès. Le médecin est invité à remplir une cause en texte libre par ligne, chaque cause indiquée sur une ligne étant « due à » celle indiquée sur la ligne suivante (relation causale attendue). Dans une **deuxième partie relative aux causes**, le médecin indique les autres états morbides, facteurs ou états physiologiques qui ont pu contribuer au décès sans être directement impliqués dans l'enchaînement conduisant au décès. Le médecin consigne aussi sur le volet médical des **informations complémentaires** notamment sur le lieu de décès, l'état de grossesse, les circonstances apparentes du décès (mort naturelle, mort subite, accident, suicide, événement pour lequel l'intention n'est pas déterminée etc.). Ces circonstances apparentes sont une nouveauté du modèle de certificat introduit en 2018, faisant suite à clarification dans la CIM 2016, elles n'étaient pas aussi détaillées, ni regroupées dans la version précédente du certificat (1997). Ces informations complémentaires suivent en partie seulement le modèle international de questionnaire adopté dans la CIM 10 – version 2016. Elles s'en écartent. Tout d'abord, le modèle international qualifie les informations complémentaires demandées de « médicales » (*frame 2 Other medical information*), ce n'est pas le cas du modèle français qui demande aussi dans cette partie d'autres questions comme le lieu institutionnel de décès, ou encore l'activité professionnelle, non requis pour coder les causes de décès, et placées au milieu du questionnaire statistique. Ensuite, la notion de « *manner of death* », qui dans le cas des morts violentes est indispensable à la classification, est traduite en français par « circonstances apparentes de décès ». Le fait que ces circonstances soient inconnues est confondu avec le fait qu'elles soient associées à une intention indéterminée (c'est-à-dire que le médecin est sûr qu'il s'agit d'une mort violente mais a des doutes sur le fait qu'il s'agisse d'un accident ou d'un suicide par exemple). Autre effet de traduction dont l'ampleur devrait être mesurée : on parle de « suicide » au lieu de « gestes auto-infligés conduisant à la mort » (*intentional self-harm*), ce qui peut référer à des intentionnalités différentes. Enfin, la précision sur comment la mort violente est arrivée (« *how did the external cause occur ?* ») n'est pas présente dans le volet français.

Le certificat des décès des moins de 1 an, certificat « infantile », vient remplacer le certificat de décès néonatal du fait de l'arrêté du 24 mai 2024. Il permet de recueillir des informations liées à la naissance des enfants décédés jusqu'à l'âge de 1 an (au lieu de 28 jours) et vient se conformer aux recommandations de l'OMS pour la rédaction des causes de décès. La partie sur les causes de décès du certificat infantile, est similaire à celle du modèle général. Les informations complémentaires sont complétées par une rubrique relative à la naissance et l'accouchement et une rubrique relative à la mère.

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 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.

En pratique, des anciennes versions des certificats restent en circulation sous format papier (versions de 1997, certificat néonatal en vigueur jusqu'au 31 décembre 2024). (<https://www.cepidc.inserm.fr/documentation/modeles-de-volets-medicaux-de-certificats-de-deces>). En 2023, encore 6% des décès étaient certifiés sur des certificats conformes à l'arrêté du 24 Décembre 1996.

Cas particulier des recherches médico-légales. 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. Le volet médical complémentaire est identique au volet médical

(initial) du certificat de décès. En pratique, ce certificat n'est que très peu renseigné par les médecins légistes (environ 600 VMC sont arrivés au CépiDc en 2022).

Cas particulier de l'institut-médico-légal de Paris. Depuis les décès de 2018, cet IML transmet au CépiDc directement des données issues de son système d'information. Celles-ci ne suivent pas le volet médical préconisé par l'OMS, il s'agit plutôt de données secondaires où le médecin légiste renseigne directement le code CIM de la cause du décès telle qu'il la définit. Elles concernent autour de 2600 décès en 2022 et permettent de résorber largement les trous de collecte ou d'information pour certains cas spécifiques notamment les homicides.

6.2 Certificat papier ou électronique

Depuis 2022 ([Décret no 2022-284 du 28 février 2022](#)), le certificat de décès doit obligatoirement être rempli sur support électronique dans les établissements de santé, sauf exceptions. Début 2025, 50 % des certificats étaient électroniques. Le site [CertDc](#) et l'application CertDc (smartphones et tablettes) permettent aux médecins ou professionnels de la santé de certifier les décès par voie électronique. La collecte de l'information sur le site suit le modèle de certificat papier. Il contient en plus des filtres et des règles de gestion relevant les incohérences. Ces règles de gestion sont détaillées dans un tableur de spécification joint au dossier, le CCTP et le SFD du marché de tierce maintenance applicative de l'application. L'application CertDc a pour MOA la DGS avec l'appui du CépiDc et pour MOE le département du système informatique de l'Inserm.

6.3 Circuit de remontée de l'information

Le médecin remplit les deux volets du certificat et scelle le volet médical pour le papier. En cas de certification papier, le certificat avec volet médical clos est transmis à la mairie. En cas de certification électronique, seule la partie administrative du certificat, imprimée ou dématérialisée, est transmise à la mairie et aux opérateurs funéraires. 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 traitements d'alignement de la base de données du CépiDc avec les données de l'Etat civil (« synchronisation », voir après, et en annexe le circuit d'un décès côté Insee).

La mairie établit l'acte de décès selon l'article 78 du Code civil³, puis rédige deux bulletins :

- l'« avis 7 » bis (B7bis), 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). Il s'agit d'un bulletin de décès à finalités administratives et statistiques.
- le « bulletin 7 » (B7), comprenant les mêmes informations sur la personne décédée, sans le nom ni prénom, généré uniquement pour le circuit papier et transmis accompagné du volet médical toujours clos à l'Agence Régionale de Santé (ARS) dans les 8 jours suivant le décès. Il s'agissait

³ « L'acte de décès sera dressé par l'officier de l'état civil de la commune où le décès a eu lieu, sur la déclaration d'un parent du défunt ou sur celle d'une personne possédant sur son état civil les renseignements les plus exacts et les plus complets qu'il sera possible. L'acte de décès peut être dressé aussitôt la déclaration effectuée et sans attendre que le certificat médical de décès prévu à l'article L. 2223-42 du code général des collectivités territoriales ait été établi par un médecin. Cette manière de faire ne présente dans la pratique aucun inconvénient sérieux, dès lors que le certificat médical de décès doit être produit pour la délivrance de l'autorisation de fermeture de cercueil. »

avant la dématérialisation des B7bis, d'une copie carbone de la partie statistique du B7bis.

Après déliassage, adresser l'avis de décès à l'INSEE et le bulletin de décès à l'ARS.

RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

7

BULLETIN DE DÉCÈS

Ce bulletin doit être adressé à l'Agence Régionale de Santé

A. IDENTIFICATION DE LA COMMUNE

Code département Code commune (1)

Libellé de la commune

N° de l'arrondissement
Paris, Lyon, Marseille

Section (2)

Type de registre Unique U Autre A → préciser

N° de l'acte N° d'ordre de l'avis de décès

B. RENSEIGNEMENTS RELATIFS AU DÉFUNT

Date du décès *jour, mois, année*

Sexe Masculin M Féminin F

Né(e) le *jour, mois, année*

A Libellé de la commune

N° de l'arrondissement Département (3)
Paris, Lyon, Marseille DOM/COM

Pays pour l'étranger

Activité principale Retraité(e) 1 En emploi 2 Au chômage 3 Autre situation 4
ly compris congé autorisé ou affecté

Profession ou ancienne profession
(3) au chômage, retraité(e) ou autre situation, indiquer l'ancienne profession

Situation professionnelle Salarié(e) de la Fonction publique (4) 1 Autre salarié(e) 2 À son compte 3

Nationalité Française 1 Étrangère 2 → préciser le pays

Domicile Numéro, voie, lieu-dit

Libellé de la commune

N° de l'arrondissement Département (3)
Paris, Lyon, Marseille DOM/COM

Pays pour l'étranger

Situation conjugale au moment du décès Marié(e) 1 Pacsé(e) 2 En concubinage ou union libre 3
Dans une autre situation (célibataire, veuf(ve), divorcé(e), etc.) 4

C. RENSEIGNEMENTS RELATIFS AU DÉCÈS

Lieu du décès Domicile du défunt ou d'une tierce personne 1 Établissement de santé public 4
(hors Ehpad, maison de retraite) Établissement de santé privé 5
Ehpad, maison de retraite 2 Autre lieu 6
Voie ou lieu public 3

Le certificat médical confidentiel de décès a-t-il été fourni ? OUI O NON N

(1) Numéro de la commune au code officiel géographique utilisé par l'Insee.
(2) À renseigner pour les communes qui possèdent plusieurs sections d'état civil.
(3) Départements métropolitains : code sur 2 positions. Pour l'outre-mer, ajouter le code en 3^{ème} position (Outre mer : Guadeloupe, Martinique, Guyane, La Réunion, Mayotte, Saint-Pierre et Miquelon, Saint-Martin, Saint-Barthélemy, Wallis-et-Futuna, Polynésie française, Nouvelle-Calédonie, Terres Australes et Antarctiques Françaises).
(4) État, territoriale, hospitalière ; non compris les entreprises publiques.

Fait le,
(date de création du bulletin)
Cachet de la mairie et signature de l'officier de l'état civil.

Vu l'avis favorable du Conseil National de l'Information Statistique, cette enquête reconnue d'intérêt général et de qualité statistique, est obligatoire, en application de la loi n°51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistique.
Visa n° 202PA009EC du Ministère de l'économie, des finances et de la relance, valable pour les années 2022 à 2026.
En application de la loi n°51-711 du 7 juin 1951 modifiée, les réponses à ce questionnaire sont protégées par le secret statistique.

Le volet médical est ouvert par l'ARS et ensuite envoyé à Insee accompagné du bulletin 7 via un prestataire de saisie et de numérisation. En cas de certification électronique, le volet médical est transmis dans les 5 minutes au CépiDc par un transfert de données depuis la plateforme CertDc.

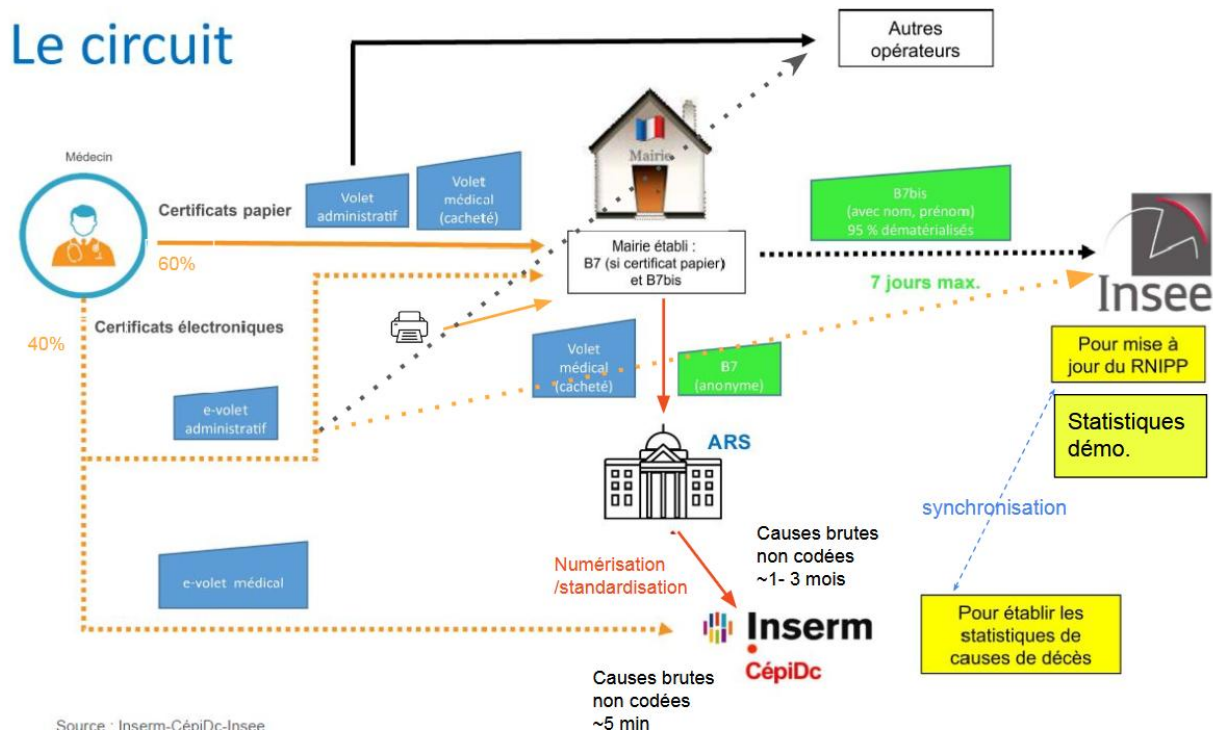


Schéma - Circuit de collecte de l'information.

6.4 Suivi de la réception des volets médicaux/contrôle de l'exhaustivité

Suivi de la transmission des certificats papiers

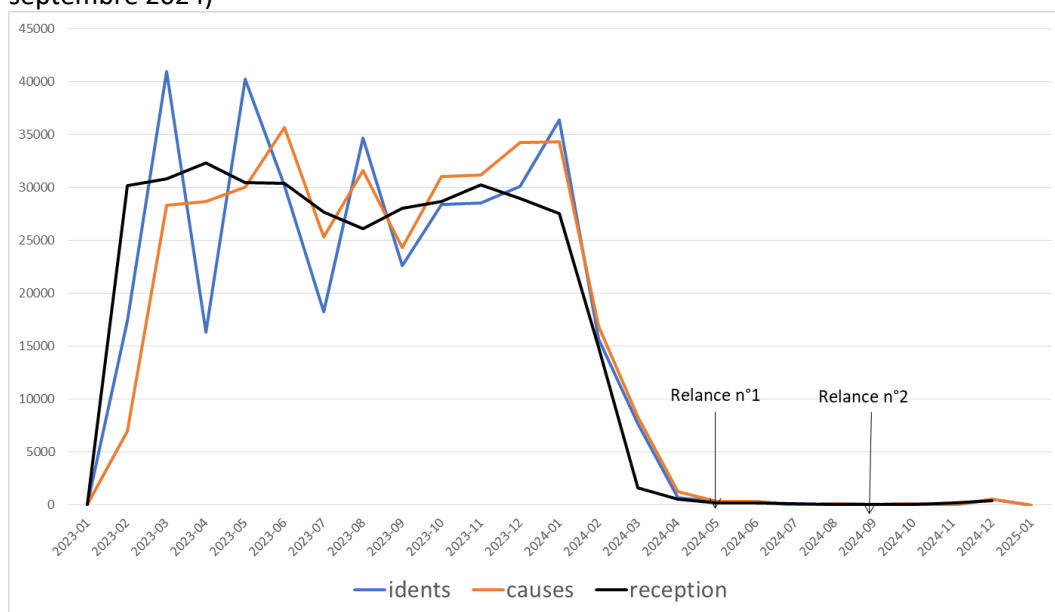
Chaque ARS doit envoyer avant le 10 de chaque mois tous les volets médicaux associés aux B7 qu'elle a reçus le mois précédent de la part des mairies. Les mairies ont elles-mêmes 8 jours pour envoyer ces documents à l'ARS. Les instructions ministérielles des mairies et des ARS définissent la mise en œuvre 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.

Suivi de la transmission des certificats électroniques

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

Suivi de collecte et campagnes de relance

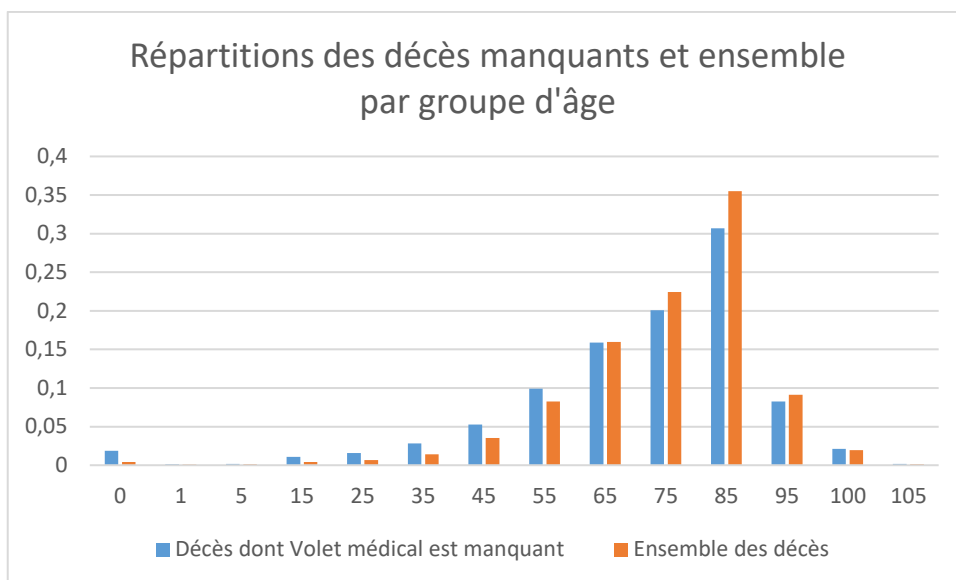
Le prestataire renvoie les données numérisées en deux temps. D'abord les données structurées et ensuite les données textuelles (causes de décès). Les envois sont quotidiens (dès qu'un pli est complètement traité, soit sur la partie structurée, soit sur la partie causes). Le CépiDc suit la réception des volets médicaux par rapport aux décès déclarés à l'Insee au fil de l'eau grâce à un tableau de bord qui compare les volumes de volets médicaux reçus aux effectifs de décès comptabilisés par l'Insee (ventilés par mois et département, à partir des fichiers de décès quotidiens publiés en mois m+1, et communes à partir des fichiers des personnes décédées publiés). Il est donc possible de cibler des relances vers les ARS pour tenter de réduire les trous de collecte en précisant les mairies pour lesquelles le trou de collecte est important. Sur les données 2023 deux campagnes de relance ont eu lieu (une en mai 2024 et une en septembre 2024)



Graphique - Réception des données relatives aux décès 2023.

Le taux de réception avant la campagne de relance de mai pour les décès de l'année 2023 était de 98,2%. Les relances ne font que dépasser de très peu ce taux, car on arrive à 98% de taux de couverture au final en enlevant les doublons et les certificats non appariables (voir après). En 2025, le CépiDc avance et compte mener les campagnes de relance à un rythme trimestriel (première relance en février 2025). A mi-mars 2025, le taux de réception pour les décès 2024 est estimé (y compris certificats non appariables qui seront supprimés ensuite) 98,2% en moyenne, 98% pour les décès de novembre et 95,9% pour les décès de décembre.

Taux de couverture final. Après traitements présentés par la suite, le taux de couverture est autour de 98% (97,99% pour 2023, 97,93% en 2022 pour 97,96% en 2021 et 97,79% en 2020). Plus de 10 000 décès identifiés à l'Etat civil et pour lesquels aucun volet médical n'a été reçu à l'Inserm sont ajoutés dans la base de données du CépiDc avec une cause initiale codée à R99 (cause inconnue) et la mention « pas de certificat » dans le texte des causes (13 000 en 2022). Il s'agit dans 53,1% des cas d'hommes, alors que ceux-ci représentent 49,8% des décès en 2022, des défunts généralement plus jeunes que dans l'ensemble des décès, y compris des bébés de moins de 1 an, caractéristiques suggérant des morts violentes.



Décès de 2022 – graphique issu du rapport de production sur les causes de décès en 2022

Ces décès « manquants » comptent pour 42% des causes initiales à R99. Le défaut de couverture est vraisemblablement lié aux circuits spécifiques aux morts violentes. En effet, en pratique lorsqu'un obstacle médico-légal est posé ou une enquête, le volet médical du certificat de décès peut ne pas être transmis à la mairie et rester dans le dossier. L'introduction du volet médical complémentaire en 2018 à remplir en fin d'enquête par le médecin légiste était censé résoudre cela. Le faible nombre de VMC reçus aujourd'hui par le CépiDc suggère un échec à ce stade de cette mesure. L'utilisation de données secondaires comme avec l'IML de Paris, donne de meilleurs résultats au prix d'information « proxy » ne collant pas aux standards de la CIM.

6.5 Saisie, numérisation, standardisation des certificats papiers et standardisation des certificats électroniques rejetés par le batch automatique de codage

Le CépiDc fait appel à un prestataire extérieur pour réceptionner les certificats de décès papier originaux, accompagnés des bulletins 7 en provenance des ARS. Ceux-ci sont numérisés et saisis, puis les images scannées et les données saisies sont envoyées au CépiDc. Le prestataire effectue aussi de premières corrections, standardisations du texte des causes sur lesquelles on revient par la suite.

Aspects sécurité dans la collecte et la remontée d'information

La sécurisation du circuit est assurée de la manière suivante :

- Le prestataire de saisi des certificats papiers (Luminess) fait signer aux personnels impliqués des clauses de confidentialité et de respect du secret professionnel qui figurent aux contrats de travail. Ces aspects sont détaillés dans le cahier des charges joint en annexe (au chapitre 3.2 du CCTP de ce marché fournis en annexe).
- Le prestataire d'hébergement des certificats électroniques se réfère au cahier des charges (CCTP fourni en annexe)
- L'authentification sur l'application de certification électronique des décès s'effectue au moyen de la carte CPS du professionnel de santé ou d'identifiant et mot de passe fournis par le référent de son établissement d'exercice ; la connexion alors établie est en https.

- Les échanges électroniques entre le CépiDc et ses partenaires : le prestataire chargé de l'hébergement de l'application de certification électronique, celui chargé de la numérisation des certificats papier, Santé publique France, l'ARS Ile de France et l'INSEE, s'appuient sur des protocoles OpenSSL pour l'authentification et le chiffrement des flux (SFTP) et sur le protocole OpenPGP pour le chiffrement des fichiers échangés avec l'INSEE (GnuPG). L'Inserm génère les clés OpenSSL et OpenPGP dont elle a la charge et obtient les certificats SSL d'un organisme reconnu (Digicert).
- Les administrateurs des serveurs de l'Inserm hébergeant les données sur les causes médicales de décès s'y connectent soit depuis le même réseau, soit via la passerelle sécurisée Wallix Bastion.

Les aspects informatiques de la sécurité du traitement sont pris en charge par le DSI de l'Inserm.

7 Description des traitements statistiques « aval »

(se rapporte aux principes n° 4, 6, 7, 8 et 12 du CBPSE)

7.1 Présentation des bases intermédiaires

a) Présentation synthétique des éventuelles bases de données intermédiaires du service statistique dont seront extraites les séries dont la labellisation est demandée (dans la mesure où les séries issues d'une même base sont a priori cohérentes sur le plan conceptuel et sur celui des données).

L'ensemble des traitements présentés ci-dessous permettent de construire la **base nationale sur les causes médicales de décès**, dont sont extraites les séries sur les effectifs de décès par cause initiale, envoyés à Eurostat en réponse au règlement européen, à l'OMS, et dont un fichier détail alimente chaque année le SNDS dans la base principale. Le dictionnaire des variables fournies au SNDS sous le format de deux tables annuelles distinctes est disponible en annexe des rapports de production (voir par exemple le rapport de production sur les données 2022 reporté en annexe).

7.2 Présentation des traitements statistiques et de la qualité associée

b) Présentation synthétique des traitements statistiques et de la qualité associée (pour chaque série, ou chaque groupe de séries traité de manière homogène) :

- **Contrôles et corrections des données** (qualité de mesure des concepts, repérage et traitement des valeurs aberrantes ou influentes, cohérence interne et externe des données, etc)

Contrôles et corrections

Sur les certificats papiers

Avant intégration en base CépiDc, des contrôles de cohérence sont réalisés sur les données structurées reçues (sexe, date de décès, date de naissance, communes de décès et de domicile, et cases cochées...) des certificats (papier uniquement, les contrôles sur les certificats électroniques se font au moment de la collecte). Ces contrôles de cohérence comparent les informations du certificat entre elles (date de décès supérieure ou égale à la date de naissance par exemple), et à celles du B7 lorsqu'il est reçu. Si une incohérence est détectée, des corrections sont réalisées soit de façon automatique, soit manuellement au travers d'une application avec interface permettant de comparer l'image du certificat à sa retranscription numérique et au B7. Certains contrôles étaient bloquants avant intégration dans la base de données, ce qui retardait le reste de la chaîne de traitements : 11% des certificats papier de 2023

avaient au moins une erreur bloquante (41 000). En 2024, un gros travail d'automatisation et de relâche du blocage à l'intégration a été réalisé. Environ 70% des cas bloqués sont débloquables (et corrigés lorsque possible) automatiquement par application de correctifs. Les principaux blocages provenaient d'absence de mention de lieu de domicile dans les volets médicaux (ceux sur le B7 ne sont pas saisis par le prestataire de saisie Luminess), d'incohérences entre B7 et volet médical sur les communes de décès, ou les dates de naissance, ou d'incohérences entre années de naissance et de décès sur le B7 mais pas sur le volet médical ou l'inverse (corrigeable). (voir annexe, note CépiDc sur proposition d'automatisation des corrections pour déblocage). Certaines corrections/ vérifications restent manuelles dans ces cas l'interface suivante est utilisée par la chargée de qualité.

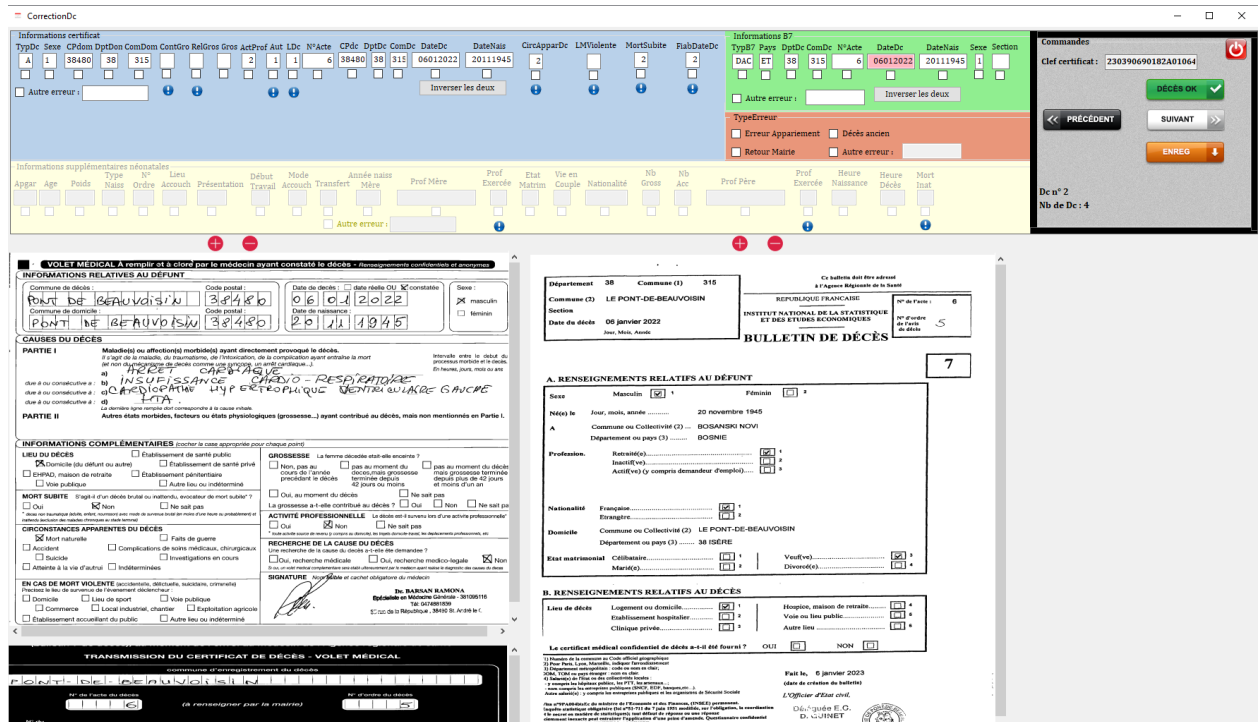


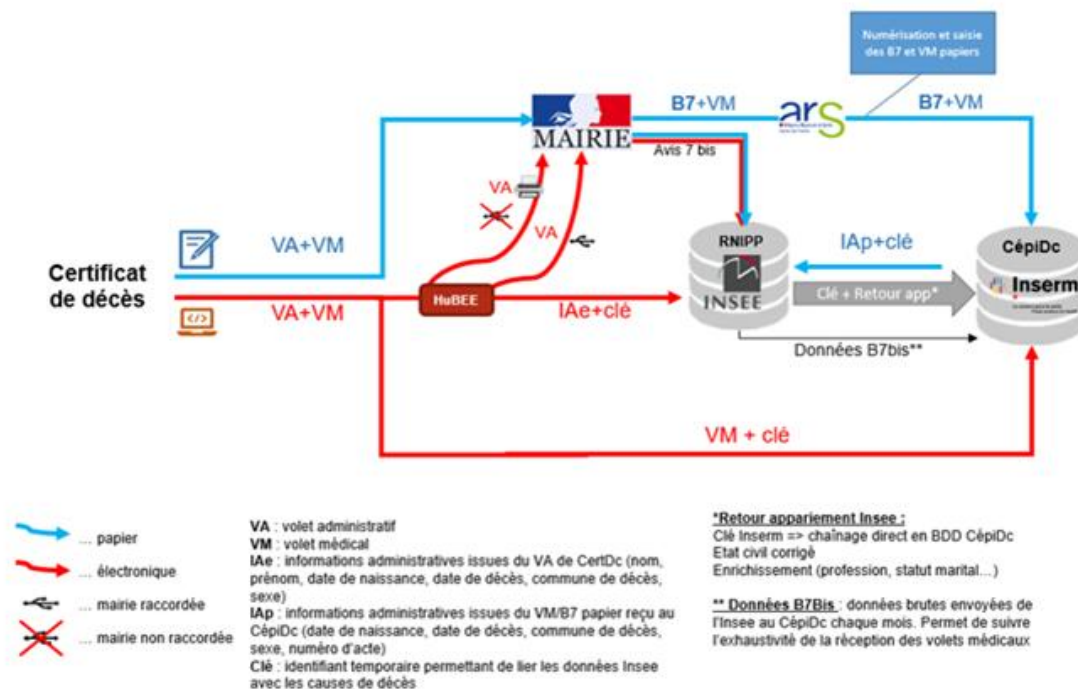
Figure - Interface de correction-numérisation

Exhaustivité – Appariement avec les données de l'Insee

- **Exhaustivité** : contrôle de l'exhaustivité, estimation du taux de couverture approché, redressements éventuels des défauts d'exhaustivité ou des valeurs manquantes (imputations, pondérations...)

Il s'agit d'un alignement avec les données de décès issues de l'Etat civil, produites et maintenues par l'Insee (Pôle Répertoire et fichiers démographiques) sur la base des variables identifiantes communes⁴ (dates de naissance et de décès, commune de décès et lorsqu'utile sexe, lieu de domicile, numéro d'acte).

⁴ Les certificats électroniques contiennent un identifiant (mais non reporté directement sur les B7bis, le champ existe mais il n'est pas alimenté automatiquement) qui n'est pas encore utilisé.



IA = informations administratives, ie variables d'appariement

Schéma – circuit de flux de données

En pratique, la méthode combine une identification au Répertoire national d'identification des personnes physiques pour les certificats électroniques, et des appariements sur les variables communes entre le volet médical du certificat de décès et le BEC avis de décès (date et lieu de décès, date de naissance, commune de domicile, sexe et numéro d'acte) sur la base des décès identifiés au RNIPP, avec relâche successive des contraintes par tours successifs. Les échanges de données permettant ces appariements sont spécifiés dans la convention Inserm-Insee.

Figure Variables identifiantes communes au certificat de décès, et à l'état civil (au B7 bis).

L'appariement est réalisé de façon mensuelle en date de réception. L'appariement est réalisé sur le *pool* de décès identifiés au RNIPP. Lorsqu'un décès a trouvé un écho, il sort du *pool* des décès à appairer (côté Inserm comme Insee) au mois suivant. Depuis septembre 2024, l'appariement mensuel ne se fait que sur critères stricts d'appariement, de façon à éviter des « faux » appariements liés uniquement à des décalages de réception d'information entre Insee et Inserm qu'il fallait « désappairer » en fin de course. En effet, l'appariement est réalisé de nouveau sur l'ensemble des décès de l'année à la clôture d'une

année de décès, en date d'événement. Les critères sont alors plus souples⁵. Cette dernière étape permet de corriger des décès appariés à tort lors des tours mensuels, et de relâcher plus de contraintes dans les tours d'appariement. A la fin, entre 1 et 2% des volets médicaux reçus au CépiDc sont supprimés des bases car ils ne sont pas appariés de manière unique à un décès Insee (dédoublonnages⁶, tests), et autour de 2% de décès correspondant à des valeurs manquantes (certificats non reçus) sont ajoutés avec une cause de décès inconnue. Ces étapes ont été mises en place pour assurer les réponses de l'Insee et de l'Inserm aux règlements communautaires sont sur le même champ.

Elles permettent ainsi de corriger et compléter les données démographiques (dates et lieux de naissance et de décès, sexe). A cette occasion, la base des causes de décès est aussi enrichie de variables socio-démographiques issues du BEC B7bis avis de décès (état matrimonial, activité, profession, situation professionnelle, nationalité, lieu institutionnel de décès déclaré à l'état civil). En effet, les différentes dématérialisations des informations ont conduit à ce que les informations statistiques du B7bis qui étaient reproduites dans le B7 n'arrivent plus au CépiDc dans un format exploitable : plus de format de B7 standard, informations non saisies....

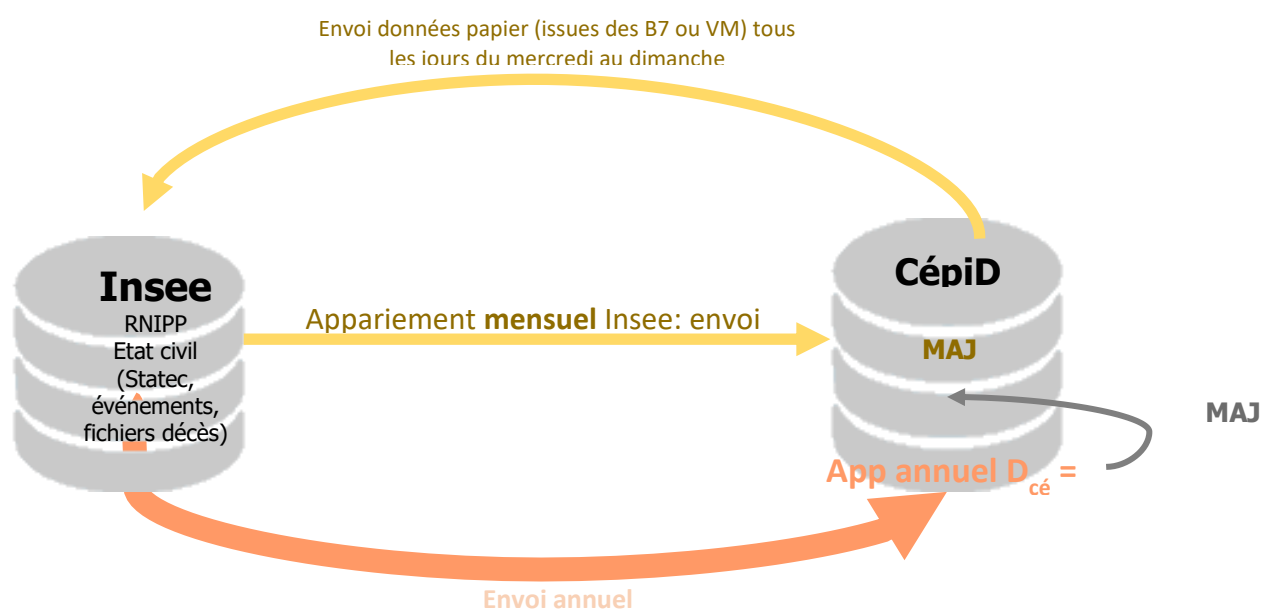


Schéma Flux Insee <-> Inserm

Codification des causes médicales de décès

- **Codifications** (nomenclatures, méthodes et qualité des codages, gestion des évolutions de nomenclature et rétropolation)

⁵ Le pôle RFD et le CépiDc mettent à jour une note descriptive du processus d'appariements [available on request].

⁶ Le « dédoublonnage » consiste, lorsque le CépiDc a reçu plusieurs certificats pour un même décès, à retenir celui apportant le plus d'informations sur les causes de décès. Il est traditionnellement réalisé à partir d'une application avec interface (arbi), mais depuis les deux dernières années il mobilise aussi les prédictions automatiques de causes (Iris et IA), ce qui en plus de règles décisionnelles automatisées, allège largement cette étape. Ainsi, si deux doublons conduisent aux mêmes informations codées, on peut en supprimer automatiquement un, sans risque.

Le CépiDc code les diagnostics inscrits sur les certificats de décès selon les règles et les rubriques de la **Classification Internationale des Maladies** (CIM8 – 8ième révision de 1968 à 1978, CIM9 - 9ième révision de 1979 à 1999 et CIM10 depuis 2000, voir Coudin Robert, 2024). L'usage de cette nomenclature est requis par le règlement européen.

La spécificité des causes de décès provient de cette information textuelle riche en termes et faiblement structurée qu'il faut transformer en un langage statistique commun. Cette information textuelle est déclarée principalement dans les Parties 1 et 2 du volet médical mais des informations médicales complémentaires sont aussi requises pour coder notamment les morts violentes.

Une première étape de standardisation du texte à l'occasion de la saisie

Les textes des causes (partie 1 et partie 2) lorsqu'ils sont saisis par le prestataire de numérisation sont standardisés (correction des fautes d'orthographe, syntaxe, séparation des entités nosologiques par des virgules et explicitations des expressions médicales). Les règles appliquées par le prestataire sont définies dans le marché (voir Annexe Règles de saisie appliquées par Luminess. Cette première standardisation facilite la codification automatique. Les certificats électroniques, rejetés par le batch automatique de codage sont aussi envoyés au prestataire de saisie et numérisation des certificats papier qui va leur appliquer ces mêmes règles. Cette étape permet d'augmenter la part des certificats électroniques codés automatiquement de 48% à 63% (comme le papier).

Objectif – coder chaque cause et déterminer la cause initiale

Coder les causes de décès dans la CIM a deux finalités distinctes : déterminer dans le texte écrit par le médecin les entités nosologiques pour leur affecter un code de la nomenclature et déterminer la cause initiale du décès, celle qui a initié le processus morbide menant au décès, le principal indicateur statistique utilisé. En 2021, un certificat non vide comprend en moyenne 3,5 codes de causes, 15 % des certificats en ont 6 ou plus (et 1 %, 11 codes ou plus). De plus, 4 145 codes différents de la CIM ont été utilisés au moins une fois, dont 2 675 codes en cause initiale.

La classification internationale des maladies – plus qu'une nomenclature statistique

« Le but de la Classification statistique internationale des maladies et des problèmes de santé connexes est de permettre l'analyse systématique, l'interprétation et la comparaison des données de mortalité et de morbidité recueillies dans différents pays ou régions et à des époques différentes (...). » Organisation mondiale de la santé, 2008, 2019, ; Rey, 2016, Coudin et Robert, 2024). Pour ce faire, la CIM décrit en complément de la nomenclature (<https://icd.who.int/browse10/2019/en>), un questionnaire de collecte de l'information et un ensemble de règles de codage, rassemblées dans un volume 2, le « guide de référence ». Ce manuel d'instruction au codage outille le codeur : une douzaine de règles appliquées sur la séquence causale selon un algorithme précis permet de déterminer la cause initiale du décès, d'une façon systématique, tout en corrigeant des erreurs ou des incohérences possibles dans la chaîne causale déclarée ou encore en privilégiant certaines pathologies à suivre car d'intérêt de santé publique (voir https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf). Ces règles s'appuient sur le modèle international du certificat.

Ces règles systématiques et précises, nées du besoin d'homogénéité pour assurer la comparabilité de la statistique dans le temps et l'espace, ont motivé l'automatisation du codage, développée en international.

Trois modes de codage dont deux automatiques et le troisième, manuel assisté en interactif

Après une production des données 2018 et 2019 dans un contexte de rattrapage, avec un plus fort recours aux méthodes automatiques, le processus de codage combine en production régulière depuis les décès

2021 trois modes de codage : le codage automatique par le système expert de règles IRIS/MUSE (63%), le codage manuel en interactif de certificats ciblés à partir du même système expert de règles mais en interactif (13%), le codage automatique IA, impliquant des réseaux de neurones profonds (RNP) entraînés sur l'historique des certificats déjà codés (25%)⁷. Les méthodes et leurs articulations sont détaillées dans Zambetta et al, 2023, 2024 ; Coudin et Robert 2024 ainsi que les rapports de production 2018-2019, 2021, et 2022. On rappelle ci-dessous uniquement leurs grands principes.

La campagne de production articule entre eux les trois modes de codage de façon à maximiser la qualité de la production dans ses différentes dimensions (dont le respect d'un calendrier de production fixé à l'avance). Les certificats rejets du batch automatique vont être affectés soit à un codage manuel assisté, soit à un codage par prédiction de réseaux de neurones profonds (RNP) en respectant les principes suivants :

- par campagne, le volume de certificats pouvant être codés manuellement est fixé à l'avance en fonction des ressources humaines disponibles dans l'équipe de codage de façon à respecter les délais de diffusion de la base statistique (autour de 90 000 certificats par année de décès) ;
- les situations pour lesquelles les prédictions des RNP sont moins vraisemblables, repérées sur la base d'un score de confiance modélisé et estimé pour chaque certificat (voir les rapports de production pour des détails sur ce score, son estimation et sa mise en œuvre), seront codées manuellement de façon ordonnée en commençant par les plus mauvaises (environ 35 000). Il en sera de même pour certains cas d'intérêt pour la santé publique qui requièrent un suivi attentif (environ 3000) : décès d'enfants en dessous d'un certain âge, morts maternelles, virus de l'immunodéficience humaine (VIH) ou syndrome d'immunodéficience acquise (SIDA) ;
- en complément, pour assurer l'entraînement et le contrôle réguliers des RNP, des échantillons tirés aléatoirement sont envoyés en reprise manuelle (environ 45 000) ;
- en fin de campagne, une partie de ces certificats codés manuellement, ainsi qu'une partie de ceux codés par batch automatique sont utilisées pour ré-entraîner les RNP et obtenir la prédiction finale des certificats codés par RNP.

Cette démarche suit des principes du code des pratiques : actualité, ponctualité, rapport coût-efficacité, etc.. Les architectures des réseaux de neurones ont été choisies pour leur simplicité. Leur entraînement et leur inférence peuvent être réalisés sur des infrastructures conventionnelles, au CépiDc. Il a été décidé de ne pas s'appuyer sur des modèles pré-entraînés, ni très complexes, pour garder entièrement le contrôle de la procédure statistique, depuis les données d'entraînement jusqu'aux modèles. On garantit le respect de la confidentialité des données, la répliquabilité et transparence de la procédure et on limite les risques de biais dans un souci d'impartialité et d'objectivité. Enfin, l'évaluation statistique qui permet de contrôler les modèles est inhérente à ces procédures. Leurs erreurs peuvent être mesurées, analysées, documentées.

⁷ Les trois modes de codage mobilisent pleinement les règles de la CIM, ce qui assure l'homogénéité du codage. C'est le cas pas construction du codage batch ou interactif par Iris/Muse et par « apprentissage » (estimation) pour le codage par RNP puisque les certificats sur lesquels les modèles apprennent ont été codés par Iris/Muse.

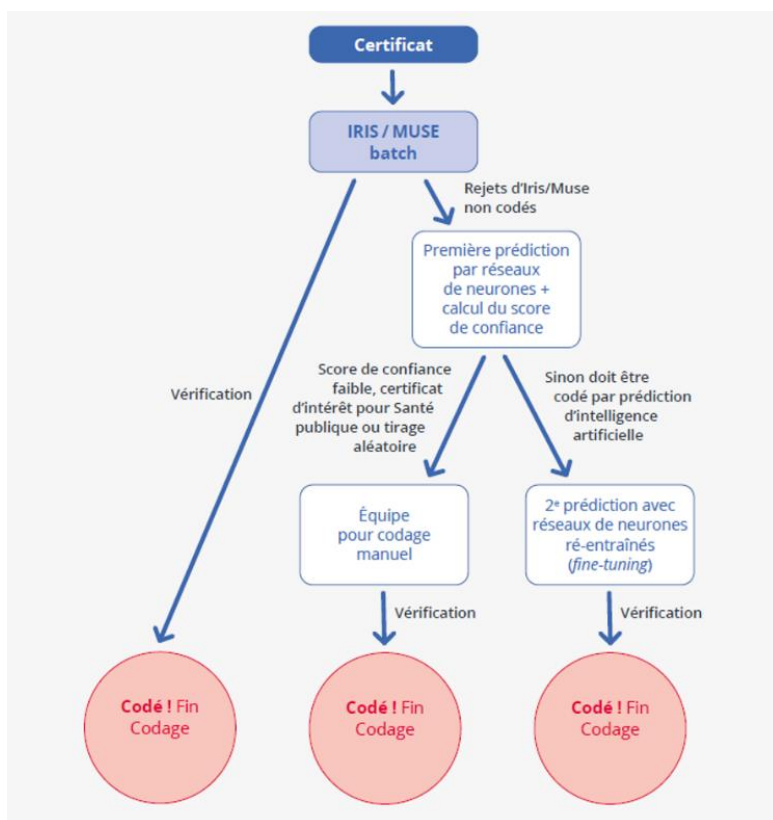


Schéma - Circuit de codage d'un certificat pendant la campagne de production

Le système expert Iris principal assure l'homogénéité du codage et la comparabilité internationale

La majorité des certificats sont codés automatiquement en appliquant le système expert de règles international Iris/Muse ([Iris](https://ec.europa.eu/eurostat/cache/metadata/EN/hlth_cdeath_simscd_fr.htm)) par batch à la réception (batch hebdomadaire). Pour les décès 2022, il s'agit de la version d'Iris 5.8.1 et de la version de Muse 2.9 [voir les metadata Causes of Death sur le site d'Eurostat pour l'historique des versions utilisées du logiciel, https://ec.europa.eu/eurostat/cache/metadata/EN/hlth_cdeath_simscd_fr.htm]. Les systèmes experts de codage des causes de décès sont utilisés en France depuis 2000, à l'occasion de la mise en œuvre de la CIM 10. Ce sont des logiciels d'aide au codage avec une interface homme-machine permettant du codage assisté, mais qui peuvent aussi être lancés en batch (Pavillon et Laurent, 2003 ; Iris Institute, 2024) Styx, puis Iris à partir de 2012, s'appuient sur le modèle international du certificat de décès et intègrent les règles du volume 2 de la CIM 10 sous forme de relations entre codes CIM, rassemblées dans des « tables de décisions » maintenues internationalement par un consortium de pays dans le cadre du Core group de l'Iris Institute (Etats-Unis, Canada, France, Italie, UK, Pays-bas, Allemagne,...). Elles suivent les mises à jour annuelles officielles de l'OMS.

Tout d'abord, il faut traduire le texte libre rédigé par les médecins en codes CIM. Cette première étape dépend de la langue. Après des standardisations du texte du certificat (synonymes, abréviations, termes non pertinents), il est mis en correspondance avec un dictionnaire contenant 157 000 expressions, pour lesquelles chaque terme est associé à un ou plusieurs codes CIM 10.

En cas de réussite, les codes de diagnostics, leurs enchaînements et les informations médicales complémentaires du modèle international (circonstances apparentes de décès, âge, sexe...) sont interprétés par la partie Muse (Multicausal and Unicausal Selection Engine, MUSE), moteur de sélection

de cause initiale et de causes multiples qui va appliquer les règles internationales d'identification de la cause initiale à partir des tables de décision. En 2022, le système Iris/Muse conclut à une cause initiale pour 63% des certificats de décès. La majorité des certificats non codés sont rejetés à la première étape (le texte ne correspond pas à un libellé unique du dictionnaire, on dit qu'il y a « choix de codes » à faire lorsque plusieurs codes s'associent au même libellé).

La majeure partie des pays européens utilisent le logiciel Iris dans le codage officiel des causes de décès (et pas uniquement les pays du consortium Core Group, les autres pays sont utilisateurs uniquement).

Codage assisté par Iris par l'équipe de codage

L'équipe de codage code systématiquement en interactif avec le logiciel Iris/Muse. Le codage manuel s'organise en fonction du niveau d'expertise dans l'équipe. Les « codeurs » corrigent, mettent en forme, simplifient le texte sans en modifier le fond, pour permettre au système expert de coder en interactif et les « nosologistes », et en dernier lieu les « experts » tranchent sur les cas délicats. Les nosologistes ont des compétences dans la terminologie médicale et la nomenclature. Ils peuvent aussi être amenés à enrichir le dictionnaire des termes rencontrés. Les experts codent les cas cliniques rares et complexes où des connaissances en physiopathologie sont nécessaires, on parle d'expertise médicale.

Codage par apprentissage profond (RNP, IA)

Dans le cadre du Plan de rattrapage et de refonte (année 2018-2019) puis en production régulière depuis 2021, un troisième mode de codage est mis en œuvre. Il repose sur des réseaux de neurones profonds entraînés sur l'historique des données passées déjà codées à transformer l'enchaînement des termes médicaux décrivant le processus morbide (séquence d'entrée) en un enchaînement de codes de la CIM et proposer une cause initiale (séquence de sortie). Il s'appuie sur les travaux d'une thèse de doctorat menée au CépiDc [Falissard et al 2020], évalués puis mis en production dans le cadre du Plan de rattrapage [Clanché et al 2023, Zambetta et al 2024], puis articulé optimalement aux autres modes de codage dans le cadre de la production régulière [rapports de production 2021, 2022].

Ces RNP sont capables de « coder » environ 100 000 certificats par jour et par machine. Contrairement au système de règles qui rejette un certificat s'il n'a pas les règles pour le coder, un RNP prédira toujours une séquence de sortie quel que soit le certificat en entrée, pour peu que son texte suive correctement la structure spécifiée. Il associera aux termes de cette séquence prédite, des probabilités qui sont par la suite être mobilisées pour estimer la qualité de la prédiction et orienter le ciblage des certificats envoyés en reprise manuelle (score).

Les RNP sont utilisés pour prédire la séquence des codes et pour déterminer la cause initiale de certains certificats. Pour déterminer la cause initiale, il y a plusieurs stratégies : soit utiliser le code directement prédit par le modèle, soit appliquer le système expert Iris-Muse sur la séquence des causes prédites par le RNP pour en déduire la cause initiale. L'approche retenue vise à retenir le meilleur des deux : elle consiste à entraîner un autre réseau de neurones capable de choisir entre les propositions possibles.

[L'article](#) du Courrier des statistique présente ces principes pour un large public [Encadré 2 notamment]. Pour les lecteurs avertis la méthode complète et son évaluation sont décrites dans l'article Zambetta et al 2024, Clanché et al. 2023, les documents de travail du CépiDc relatifs aux campagnes de production 2018/2019, 2021, et 2022.

Vérification de codage

Quel que soit le mode de codage utilisé pour les coder, les certificats dont le codage apparaît comme « potentiellement incohérent » (d'après des repérages déterministes d'association improbable de codes ou de codes et d'âge, sexe...) ou quand ces certificats concernent des décès à fort enjeu en termes de santé publique, ils sont vérifiés manuellement, voir corrigés. Il peut s'agir de vérifications liées à des incohérences manifestes (par exemple entre le sexe ou l'âge et la cause de décès, vérifications dites de

« cohérence »), des vérifications liées à des anomalies identifiées du système expert, ou encore liées à des évolutions de règles de codage pour lesquels il est nécessaire de vérifier leur bonne application.

A titre d'exemple, en 2022, environ 11 000 certificats ont ainsi été vérifiés, conduisant à une modification de la cause initiale (au niveau le plus fin de la CIM) dans un tiers des cas. Le scope et les résultats de cette phase sont décrits dans le rapport de production de l'année. En fin de campagne, une évaluation de la pertinence des vérifications est réalisée afin d'optimiser ce travail pour la campagne suivante (voir rapport de production 2022).

Géographie

Géographie (qualité de l'information brute et le cas échéant de la codification, taux de couverture de l'organisme aux différents niveaux géographiques, gestion des évolutions de la géographie ou de la nomenclature géographique et rétropolations)

Comme indiqué plus haut, les variables géographiques (lieux de résidence, lieux de décès) suivent le COG et ses évolutions dans le recueil de l'information et sont mises en cohérence au travers de la synchronisation avec ces mêmes informations déclarées à l'Etat civil et corrigées par l'Insee dans le cadre de l'établissement des statistiques démographiques.

Méthodes de calcul des indicateurs, corrections des variations saisonnières...

Méthode de calcul des indicateurs, sources externes utilisées le cas échéant, Corrections de variations saisonnières, pour jours ouvrables (préciser la méthode)

Voir plus haut plus les méthodes de calcul des taux standardisés de mortalité par cause. Il n'y a pas de correction des variations saisonnières, ou pour jours ouvrables.

Données provisoires – *fast estimates*

En cas de diffusion de **données provisoires** : s'agit de données provisoires incomplètes (caractéristiques et limites à préciser) ou bien complétées (provisoirement) par le biais de méthodes d'extrapolation ou de prévisions (méthode et qualité à décrire) ; ampleur et gestion des révisions entre données provisoires et définitives.

Les données provisoires ne sont pour le moment pas diffusées. Depuis la diffusion des causes de décès de l'année 2021, des estimations (*fast estimates*) issues d'un traitement automatique l'ensemble du champ (Iris+ prédiction RNP sur l'ensemble, sans reprise manuelle) concernant de l'année N+1 sont présentées dans un encadré de l'Études et Résultats marronnier accompagnant la diffusion des causes de décès de l'année N. On se concentre sur les grandes causes de décès en effectifs et sur lesquelles les prédictions par RNP sont généralement fiables (effectifs arrondis au 1000er, taux à l'unité, exemple tiré de l'ER sur les décès 2022 ci-dessous).

Encadré 2 Premières estimations par cause de décès en 2023

En 2023, autour de 638 000 décès ont eu lieu en France y compris des décès de non-résidents, selon l'estimation encore provisoire mais révisée du bilan démographique de l'Insee (Papon, 2024) et dans Insee (2024). Cet effectif est en baisse par rapport à 2022 (675 000). Une première exploitation des certificats de décès de 2023 reçus en septembre 2024 au CépiDc et dont les causes ont été codées dans la nomenclature CIM-10 de façon majoritairement automatique (par système de règles ou par prédiction d'intelligence artificielle, le codage manuel venant juste de commencer) permet de dresser des pistes d'évolutions de la mortalité pour certaines grandes catégories de causes de décès. Une approche similaire avait également été réalisée l'an passé pour les données de 2022 (Cadillac, et al., 2023). Il est ainsi possible, pour 2022, de revenir sur les écarts entre les premières estimations (résultats provisoires) et les résultats définitifs, après campagne affinée de codage, tels que présentés dans cette publication.

Les écarts, en effectifs comme en taux, entre les résultats 2022 provisoires et les résultats 2022 définitifs sont faibles. Ils égalent ou dépassent rarement 1 000 décès pour les effectifs de chaque grande catégorie, et l'unité pour les taux. Ces écarts incitent cependant à ne pas interpréter des évolutions d'effectifs inférieures à 1 000 décès et de taux inférieurs à une unité (tableau).

En 2023, pour l'ensemble des causes de décès publiées ici, la mortalité serait en baisse. La baisse la plus importante concernerait le Covid-19, avec environ 27 000 décès de moins qu'en 2022 et un taux de mortalité estimé de 18, soit 37 décès de moins qu'en 2022 pour 100 000 habitants. La mortalité due aux tumeurs baisserait légèrement et le nombre de décès se stabiliserait (écart de taux de -4 décès pour 100 000 habitants, même nombre de décès). Les décès dus aux maladies cardio-vasculaires seraient quasiment stables (-1 000 décès) et la mortalité diminuerait (-5 décès pour 100 000 habitants) du fait de hausses des pathologies cardiaques et des maladies cérébrovasculaires.

La mortalité due aux maladies de l'appareil respiratoire augmenterait légèrement (+2 décès pour 100 000 habitants et +2 000 décès). Elle dépasserait alors légèrement son niveau de 2019, avant la crise due au Covid-19.

Effectifs et taux standardisés de décès de 2021 et 2022 et estimations de 2022 et 2023

	Effectifs				Taux standardisés de décès			
	2021	2022 (p)	2022	2023 (p)	2021	2022 (p)	2022	2023 (p)
Tumeurs	170 000	172 000	172 000	172 000	243	242	242	238
Tumeurs du colon, rectum, anus	17 000	17 000	17 000	17 000	24	23	23	23
Tumeurs du pancréas	13 000	13 000	13 000	13 000	18	18	18	18
Tumeurs du poumon	30 000	31 000	31 000	31 000	45	45	45	45
Tumeurs du sein	13 000	13 000	13 000	13 000	15	15	16	15
Maladies endocriniennes, nutritionnelles et métaboliques	24 000	26 000	26 000	25 000	31	33	32	32
Diabète sucré	12 000	12 000	12 000	12 000	17	16	16	16
Autres maladies endocriniennes, nutritionnelles et métaboliques	11 000	13 000	13 000	13 000	14	16	16	16
Troubles mentaux et du comportement	24 000	26 000	26 000	25 000	30	31	31	30
Maladies du système nerveux et des organes des sens	36 000	38 000	38 000	37 000	46	48	48	46
Maladies cardio-vasculaires	138 000	141 000	140 000	138 000	178	179	178	172
Pathologies cardiaques	82 000	83 000	83 000	81 000	107	106	107	103
Maladies cérébrovasculaires	31 000	32 000	32 000	31 000	40	40	40	38
Autres maladies de l'appareil circulatoire	25 000	26 000	25 000	26 000	31	32	32	31
Maladies de l'appareil respiratoire	36 000	45 000	45 000	47 000	49	60	59	61
Grippe et pneumonie	10 000	16 000	16 000	17 000	14	20	20	21
Maladies chroniques des voies respiratoires inférieures	10 000	12 000	11 000	12 000	14	16	16	16
Autres maladies de l'appareil respiratoire	16 000	18 000	18 000	18 000	21	24	23	23
Maladies de l'appareil digestif	26 000	27 000	27 000	27 000	35	37	37	36
Maladies de l'appareil génito-urinaire	13 000	13 000	13 000	13 000	17	17	18	17
Covid-19	61 000	41 000	41 000	14 000	83	54	55	18

p : estimations.

Notes > Les effectifs sont arrondis au millier et les taux à l'unité, au vu des incertitudes provenant des estimations.

Le caractère provisoire des chiffres ne permet pas encore de publier un taux de mortalité toutes causes confondues sur ce champ.

Lecture > On estime 14 000 décès dus au Covid-19 en 2023, avec un taux standardisé de mortalité estimé de 18 pour 100 000 habitants.

Champ > Personnes décédées en France et résidant en France.

Source > Inserm-CépiDc, traitement DREES.

> Études et Résultats n° 1312 © DREES

Autres travaux de validation - évaluation de la campagne annuelle de production

(robustesse...), dont validation des résultats (plausibilité des évolutions ou écarts entre catégories...)

Introduire des RNP dans une production statistique nécessite de les « superviser », c'est-à-dire d'évaluer régulièrement leur performance en comparant leurs résultats à ceux des autres méthodes et de les réestimer régulièrement pour prendre en compte les nouveautés de vocabulaire, de description des pathologies ou les modifications dans les règles de codage. La performance des RNPs s'évalue en comparant la séquence de sortie prédite par le modèle avec la séquence véritablement observée sur un jeu de données déjà codées qui n'a pas été utilisé dans l'entraînement. Ce principe, essentiel en apprentissage statistique, permet de pallier le risque de sur-apprentissage de ces modèles qui par nature comprennent plus de paramètres à estimer qu'il n'y a d'observations dans la base d'entraînement.

En pratique depuis le rattrapage, le CépiDc évalue chaque année la « qualité » d'une campagne de production, et l'apport de chacune de ses étapes, en estimant l'écart qu'il y a entre les statistiques issues de la campagne de l'année et celles à auxquelles aurait conduit une campagne traditionnelle mobilisant uniquement le codage automatique ou assisté par Iris/Muse. Sur un jeu de test de référence n'entrant pas dans les entraînements des RNP et représentatif de la distribution des causes de décès en population générale, on peut estimer la cohérence de la campagne de codage d'une année donnée en comparaison à une campagne traditionnelle (utilisant uniquement le codage automatique par le système expert et le codage manuel) en simulant l'articulation des modes de codage tel qu'il a été réalisé lors de la production. Ainsi, en 2022 en prenant en compte le fait que 63 % des certificats ont été codés par batch et que pour ces certificats, le codage par rapport à une campagne classique ne change pas (la cohérence est parfaite), dans 95,9 % des cas, la cause initiale issue de la campagne à trois modes de codage est la même au niveau le plus fin de la CIM que celle que l'on aurait obtenue par une campagne classique de codage, et dans 97,4 % des cas, dans la même rubrique de la shortlist européenne, et ceci avec uniquement 13 % de codage manuel et un respect des délais impartis. Ces résultats sont présentés et expliqués dans les rapports de production annuels qui contiennent aussi des indicateurs par type de cause (précision, rappel, F-mesure, effectifs estimés). Ils mettent en avant les risques de rupture de série (et évaluent leur ampleur) associés à l'introduction des RNP articulés aux autres modes de codage pour certaines catégories de causes. Par exemple (chiffres issus de rapport de production 2022). On s'attend à un écart de -4% sur les effectifs des décès dus à des maladies du sang ou hématopoïétiques et +2% pour les démences (chiffre important à signaler dans le contexte de population âgée plus nombreuse), alors que pour les décès dus à des tumeurs les écarts sont quasiment nuls quelle que soit le type de tumeur.

7.3 Compléments méthodologiques relatifs à des enjeux ou séries spécifiques

c) Compléments méthodologiques relatifs à des enjeux ou séries spécifiques Pour des séries et / ou des enjeux particuliers (sensibilité, nouveauté, refonte, choc important, problèmes de qualité, proximité avec des séries de la statistique publique ou labellisées par l'ASP...) : travaux de validation, écarts (en niveau ou évolution) avec des séries de référence d'autres producteurs du SSP ou labellisées etc.

Utiliser des réseaux de neurones en production de statistique officielle est plutôt innovant. Cette démarche a donné lieu à divers types de communication, support : rapports de production, documents de travail, un article académique Zambetta et al 2024 en anglais, de la documentation spécifique accompagnant le rapport qualité « metadata » mise en ligne sur la page de metadata d'Eurostat en anglais aussi, et ainsi partagés auprès des homologues internationaux. Ces documents développent la stratégie de validation et fournissent des mesures des écarts induits. Voir https://ec.europa.eu/eurostat/cache/metadata/EN/hlth_cdeath_simsd_fr.htm et la note tout en bas de la page web.

D'autres sujets plus spécifiques donnent aussi lieu à des évaluations (*a posteriori*). Récemment, notamment, l'effet de l'introduction du nouveau modèle de certificat incluant les cases cochées (et la case cochée suicide) a été étudié dans le cadre d'une thèse de médecine (O. Pierre 2023). S'appuyant sur deux méthodes distinctes (évolution des « termes » associés au suicide déclaré sur le certificat, et économétrie via matching sur score de propensité), il est évalué que la hausse apparente du taux de suicide entre 2017 et 2020 peut entièrement s'expliquer par l'introduction du nouveau certificat, ainsi que l'arrivée des données de l'IML de Paris. Cette hausse de l'ordre de 10% est du même ordre de grandeur que l'estimation en 2006 de la sous-estimation du suicide [Aouba, Péquignot, Camelin, Jouglu, 2011]. Ces travaux ont été présentés à l'Observatoire national du suicide par une note de la Drees. La non-interprétabilité de la tendance entre 2017 et 2020 est rappelée et discutée dans les publications marronnières.

8 Existence et gestion de ruptures de séries passées ou à venir

(se rapporte aux principes n° 6, 7, 8, 12 et 14 du CBPSE)

8.1 Evolutions passées et à venir

a) *Évolutions passées ou prévues, à court ou moyen terme, du processus de production (évolution du champ, modifications importantes des sources ou de la réglementation, changements de méthodes ou de calendrier...). Traitement de l'impact de ces évolutions sur les ruptures de séries : rétroplation, documentation et information des utilisateurs.*

Il existe de nombreuses ruptures sur les séries des causes de décès. Et d'ailleurs, l'enjeu de les détecter, les mesurer a conduit à des publications de méthodologie appliquées (repérage de sauts automatiques dans les séries). Voir notamment Rey et al 2011.

Changements de version de la CIM. Il s'agit de révisions profondes de la CIM susceptibles de modifier l'arborescence de la classification et les règles de codage. Ces changements sont connus et bien identifiés avec des effets (et des mesures) en international. Des « bridge coding » ont été réalisés pour en mesurer l'impact sur les statistiques⁸, notamment 2000 est l'année d'introduction de la CIM 10, qui a engendré des ruptures pour de nombreuses catégories.

« Evolutions » de la CIM. Changements intervenant régulièrement pour introduire de nouveaux codes ou en faire évoluer d'autres, sans changer de version de la CIM. Ces évolutions sont prévues depuis la 10^{ème} révision de la CIM. Les impacts sont généralement ciblés sur certaines pathologies. Les effets des ruptures induites sont mesurés sur une base statistique : la distribution de la mortalité est comparée sur plusieurs années par cause initiale des décès. Les ruptures de série sont quantifiées et vérifiées côté statistique et spécialistes de la classification. Depuis l'introduction des rapports de production, ces modifications sont systématiquement documentées. A noter que les évolutions de la CIM concernent tous les pays, et que ces ruptures donc doivent se retrouver dans les comparaisons internationales. Les principales évolutions à noter ont eu lieu en 2016 (changement du questionnaire de collectes+ règles), et 2019 (Covid).

Codage automatique par système expert. Changements dus à la mise en place ou l'usage plus étendu d'un logiciel de codage automatique. Des études sont menées lorsque les changements sont conséquents (ex. changement du moteur de codage). Elles ne conduisent pas forcément à publication. La liste des versions /spécifications utilisées du logiciel Iris-Muse est disponible dans les metadata https://ec.europa.eu/eurostat/cache/metadata/EN/hlth_cdeath_simsd_fr.htm. A noter que tous les pays utilisant IRIS /muse utilisent les mêmes versions / spécifications de l'année, lesquelles suivent les évolutions annuelles de la CIM.

Changements de modèle du certificat/questionnaire. Souvent en lien avec une « évolution » de la CIM mais décalés au temps d'implémentation nationale, ces changements peuvent être diffus car les certificats papiers sont distribués aux médecins au fur et à mesure de l'épuisement des stocks existants. La certification électronique permet d'éviter cet effet diffus. Il n'existe pas de moyen simple de connaître l'impact de ces changements sur les séries. Plusieurs études ont été menées sur des sujets variés et avec des méthodes d'évaluation variées : Comparaison de la certification électronique aux données issues des

⁸ Pavillon G, Boileau J, Renaud G, Lefèvre H, Jouglà É. Conséquences des changements de codage des causes médicales de décès sur les données nationales de mortalité en France, à partir de l'année 2000. *Bul Epidémiol Hebd.* 2005;4:13–6

certificats papier⁹. Cette étude retrouve que les certificats électroniques sont plus complets et plus cohérents. Thèse d'O. Pierre sur l'impact du nouveau modèle de certificat sur l'évolution du taux de suicide déjà mentionnée précédemment.

Nouveau mode de codage, via le codage par réseau de neurones. La rupture en 2018, 2019, puis 2021 etc est évaluée, mesurée, documentée sur des supports de documentations variés avec pour objectif de toucher largement son public, mentionnée aussi dans les notes de lecture des Grandes tendances.

Modifications conséquentes des traitements relatifs à l'exhaustivité. Synchronisation avec l'Insee. En effet, qu'elles proviennent d'effet sur les variables d'appariement, ou sur la qualification des tours d'appariements (critères plus ou moins stricts), toute modification peut avoir des effets ensuite sur le taux de couverture « estimé » et le nombre de certificats « ajoutés » en non-réponse (R99) pour atteindre l'exhaustivité.

Utilisation de données secondaires. Point lié à celui du dessus. L'utilisation des données secondaires de l'IML de Paris à partir de 2018 permet de capter plus de morts violentes (en particulier des homicides et des suicides).

Les ruptures de séries étaient dans le passé généralement étudiées. Les résultats des études donnant lieu à publication scientifique. Cela a moins été le cas pour la rupture spécifique à 2016, période où le CépiDc n'avait pas de moyens (2020), combiné à la crise de 2020 et la priorisation sur le rattrapage.

Peut-être que ce qui manque c'est un **recensement, une synthèse et une vulgarisation** de l'ensemble de ces ruptures dans la documentation plus grand public mise à disposition sur le site du CépiDc ou accompagnant la diffusion de la base au SNDS. C'est un travail en cours (voir le tableau Etat des lieux des ruptures de série, en annexe de tableaux, dans sa version provisoire qui n'est pas encore mise en ligne).

Les prochaines ruptures de série à venir sont :

Certification par les infirmiers. Depuis un an, via une expérimentation sur certaines régions, le personnel infirmier est habilité à certifier des décès si un médecin n'est pas disponible dans des cas simples (EHPAD, domicile, hors morts violentes, personne majeure). Cette expérimentation est généralisée à partir du 26 avril 2025. Le CépiDc a mesuré l'impact de cette extension des certificateurs sur le taux de codage automatique sur les certificats électroniques (car le numéro RPPS permet de savoir si le certificateur est médecin ou infirmier). Toutes choses égales par ailleurs (matching sur score de propensité) le taux de codification automatique des certificats de décès remplis par un infirmier était 16 points plus bas (sur mai à fin octobre) et 12 points plus bas (sur mai à fin décembre) que celui des certificats de décès remplis par un médecin dans une situation comparable de certification. Le CépiDc a notifié la DGS¹⁰ et des échanges sont en cours sur la formation des infirmiers. Pour le moment, il n'y avait pas encore assez de cas pour aller plus loin dans l'évaluation, mais il est possible que ce phénomène affecte aussi les causes et les processus morbides déclarés (moins pertinents et synthétiques ce qui conduirait à la baisse du codage automatique). Sur les certificats électroniques, c'est tout-à-fait évaluable.

Nouvelle révision de la CIM – CIM 11. L'OMS a adopté en janvier 2022 la 11e version de la Classification internationale des maladies (CIM-11), qui doit à terme remplacer la 10e version. Aucun calendrier d'implémentation n'est à ce jour défini/imposé. Les pays européens utilisateurs d'Iris les plus avancés

9 Lefeuvre D, Pavillon G, Jouglu E, Aouba A, Rey G. Évaluation de la qualité des certificats de décès en France : l'apport de la certification électronique. *Bul Epidémiol Hebd* 2013 7;57-60

¹⁰ La note est disponible sur demande

annoncent 2028 comme année de première implémentation, lorsque le logiciel Iris 6 pour coder en CIM 11 sera disponible. Ce sera une profonde rupture de série, car la structure et les objectifs de la classification changent. La CIM 11 a un objectif d'interopérabilité et de réutilisation qui dépasse celui de la statistique de santé. Les pays travaillent à la transition de concert. Eurostat organise une task force pour traduire en CIM 11 la shortlist européenne (de façon à limiter les ruptures de séries). Les pays du Core group Iris travaillent à l'actualisation du logiciel Iris. L'ONS au Royaume-Uni, pays le plus avancé sur la stratégie d'évaluation, a programmé une campagne de double-codage sur 18 mois à partir de mi-2026. Ils sont ouverts à partager avec les autres pays européens les données (non identifiantes) pour les mesures d'évaluation. Pour le CépiDc, des premiers travaux sur le dictionnaire (alignement à la CIM 11, alignement entre CIM 10 et CIM 11 utilisant des modèles d'IA pré-entraînés) sont en cours. Suivront des tests de la version 6 d'Iris/Muse et de Doris l'outil de détermination de la cause initiale produit par l'OMS. A noter que les réseaux de neurones utilisés actuellement ayant « appris » sur des certificats codés en CIM 10 ne seront pas capables de coder directement en CIM 11 sans adaptation particulière. L'objectif pour le CépiDc est plutôt de profiter de cette occasion pour augmenter largement le taux de codage automatique par le logiciel Iris (voire Doris). Les réseaux de neurones utilisés actuellement permettent en revanche d'effectuer à moindre coût les campagnes de double codage.

L'OMS a développé en parallèle de la CIM 11, aussi des outils de codage, dont un DORIS qui va jusqu'à déterminer la cause initiale (et joue le même rôle qu'IRIS). Se posera la question de son utilisation ou non.

Changement de questionnaire ? S'il est recommandé d'adapter le questionnaire/modèle de certificat au modèle international, cela induira une rupture de série sur les causes des morts violentes. Il n'y a pas encore eu de réflexion sur la façon d'en évaluer l'impact.

8.2 *Processus de revue périodique des choix et spécifications méthodologiques.*

La mise en place de la gouvernance de la statistique sur les causes de décès rend possible une validation annuelle des choix méthodologiques. Mais instaurer une obligation de revue méthodologique plus profonde à intervalle régulier devant un comité expert extérieur (comme le Cnis ?) pourrait être bénéfique.

9 Gouvernance et rapports d'audit récents

(se rapporte aux principes n° 11 et 13 du CBPSE)

- a) *Rôle des éventuelles instances de suivi : comité scientifique, de concertation, d'orientation, de pilotage, de suivi, etc. Composition des différents comités.*
- b) *Audits externes réalisés.*

9.1 *Mission interIG sur les statistiques de décès et mission de rattrapage*

En tant qu'« autorité nationale » productrice de statistique officielle, le CépiDc doit rendre compte devant l'Autorité de la statistique publique de la qualité de sa production statistique suivant les critères du code des bonnes pratiques en matière de statistique européenne et se plier aux règlements relatifs. L'ASP avait labellisé les séries statistiques des causes de décès pour les données produites entre 2017 et 2022 avec des recommandations / des axes d'effort dont en particulier un engagement signé du Président directeur général de l'Inserm à publier les données dans les délais réglementaires et une attention particulière aux ressources humaines notamment car une vague de départs à la retraite s'annonçait. Pourtant les délais de diffusion des données ont empiré, et les ressources du CépiDc diminuées de 2018 à 2020, puis en 2021 avec la séparation de l'unité en deux unités distinctes aux missions redéfinies (CépiDc et France Cohortes). Les données de décès 2016 ont été publiées avec 6 mois de retard en 2019. Puis il n'y a plus de diffusion jusqu'à septembre 2021 (décès 2017, 23 mois de retard).

En 2021, une mission inter-inspection Igésr-Igas-IG Insee sur les statistiques de décès a soulevé toutes une série de problèmes dans la chaîne de production, amplifiés dans la période récente, et qui conduisaient à allonger encore le temps de traitement des données. Fin 2021, l'Autorité de la statistique

publique appelle au lancement immédiat d'une mission de rattrapage avec moyens humains, informatiques et financiers au CépiDc et à la Drees dédiés. Pour répondre à ces problèmes structurels, l'Inserm inscrit dans les objectifs de rattrapage et de maintien pérenne des missions du CépiDc dans le Contrat d'objectifs, de moyens et de performance 2021-2025 (COMP), la « redynamisation du CépiDc ». L'Inserm consacre 800keuros du COMP au CépiDc et s'engage à financer 15 postes sur son plafond actuel d'emplois. Le MESR abonde d'un financement supplémentaire de 1,5Meuros (salaires+fonctionnement) + 500keuros pour refondre le système d'information.

La mission interministérielle de rattrapage et de refonte des statistiques de décès est lancée tout début 2022. Elle est dirigée par un directeur attaché auprès du DREES (François Clanché) qui travaille de concert avec le puis la directrice du CépiDc (Grégoire Rey puis Elise Coudin). Les financements du projet de rattrapage permettent d'embaucher des codeurs, un statisticien de production et de constituer une petite équipe de datascience mais ces financements ne sont pas pérennes. Fin 2022, les données sur les causes de décès 2020 sont produites et diffusées dans les temps impartis. Puis les données 2018 et 2019 sont produites en mobilisant en partie une nouvelle méthode de codage à partir de réseaux de neurones profonds. Ce nouveau mode de codage entre dans la production courante à partir des données 2021 diffusées en décembre 2022, et permet d'accélérer le calendrier de diffusion avec une diffusion des données de 2022 en septembre 2024.

L'unité de service CépiDc doit être renouvelée en 2025. Elle a déposé une demande de renouvellement et sera visitée par la mission accompagnement de l'Inserm + un cadre de la statistique publique (cf section 2) en avril 2025.

9.2 Gouvernance

Depuis octobre 2024 et faisant suite au Plan de rattrapage et de refonte de la statistique sur les causes de décès, les tutelles de l'Inserm et l'Insee ont demandé à l'Inserm de mettre en place un comité stratégique assurant la gouvernance du CépiDc (voir lettre reportée en annexe). Ce comité a pour missions pérennes :

- de vérifier si les moyens du CépiDc sont quantitativement et qualitativement en adéquation avec les missions qui lui sont confiées
- de valider les choix d'organisation et de méthode du CépiDc et, le cas échéant, lui demander de les faire évoluer ;
- de conseiller et d'assister le CépiDc et les autres acteurs de la chaîne de production des données pour respecter la qualité de la statistique publique ;
- d'organiser les échanges entre tous les partenaires responsables de la production de la source (certificats de décès) et de sa mise à disposition.

Cette instance, présidée par le directeur de la DREES, rassemble des représentants :

- de la direction générale de la recherche et de l'innovation et de la direction générale de la santé et, tutelles de l'Inserm ;
- de l'Institut national de la statistique et des études économiques, autorité statistique nationale, garant in fine de la qualité des statistiques produites sous règlement européen ;
- de la direction de la recherche, des études, de l'évaluation et des statistiques, responsable de la statistique publique dans le domaine de la santé ;
- de l'Inserm, auquel la mission de production et de diffusion de la statistique sur les causes de décès est confiée ;

- des autres parties prenantes du processus de production et des principaux utilisateurs institutionnels des données produites : Santé publique France, les agences régionales de santé, la Caisse nationale d'assurance maladie.

L'existence et les missions de ce comité sont inscrites dans le règlement intérieur de l'unité. Le comité stratégique se réunit au moins tous les semestres. Cette gouvernance est inscrite dans le règlement intérieur du CépiDc.

10 Lettre d'engagement sur le calendrier de publication et de diffusion des séries

*Explication, le cas échéant, des règles et modalités d'embargo.
(se rapporte au principe n° 6 du CBPSE)*

La lettre d'engagement signée par le PDG de l'Inserm se trouve en Annexe 22 (la dernière).

Depuis la mise en place des publications marronnières accompagnant la diffusion des données 2020, 2021, 2022, sur deux supports complémentaires (BEH et Etudes et Résultats), la diffusion des données est systématiquement accompagnée d'une première analyse co-écrite par le CépiDc, la Drees et Santé publique France. La publication sur les deux supports est simultanée. Elle facilite et oblige à une cohérence en termes de politique de diffusion et au respect du code (envoi sous embargo court et restreint à des personnes identifiées et prévues à l'avance, calendrier prévu et annoncé à l'avance). La publication est accompagnée d'un communiqué de presse commun aux trois institutions. Les sites internet sont prêts un peu en avance à diffuser les résultats en cas de rupture d'embargo (ce qui n'est pas arrivé). A titre d'exemple, voir en annexe la note préparant la diffusion des publications sur 2022 et les étapes au calendrier fixé. C'était aussi le cas l'année passée et ce qui est prévu pour l'année prochaine.

11 Modalités de présentation et de diffusion des séries

(se rapporte aux principes n° 1, 4, 6, 11, 13 et 15 du CBPSE) Ce point énumère des préconisations que devra suivre le service lors de la diffusion des données labellisées, afin notamment de permettre à un utilisateur externe de ces données non spécialiste du champ d'accéder à toutes les informations utiles pour les interpréter correctement (et le cas échéant les rapprocher d'autres données). Le service doit fournir les éléments permettant de juger de la conformité de ses pratiques de diffusion à ces principes. Le cas échéant, il indiquera les raisons de certains écarts, ou un calendrier de résorption de ces écarts.

11.1 Accessibilité

rapport, site internet, facilité à trouver les données labellisées sur le site, données téléchargeables avec les outils usuels ;

Une page du site internet oriente l'utilisateur vers le type de données répondant le mieux à ses besoins : <https://www.cepiddc.inserm.fr/donnees-et-publications/interroger-les-donnees-de-mortalite>

La base individuelle sur les causes médicales de décès est disponible pour les chercheurs au SNDS. Les séries longues d'effectifs sont disponibles et téléchargeables sur le site du CépiDc au niveau national sur la page [Grandes tendances](#) du site internet du CépiDc pour les données depuis 2015 et sur [l'Opendata](#) depuis 1979 avec des ventilations plus fines. Les données sont téléchargeables pour faciliter leur réutilisation.

Sur le site internet, ces données sont accompagnées à la fois des liens vers les publications marronnières qui accompagnent la mise à disposition des données annuelles et d'une fiche méthodologique qui explicite les définitions et les calculs des indicateurs mis à disposition. Pour les utilisateurs plus avertis, les

mises à disposition de données s'accompagnent depuis le millésime 2018 des rapports de productions des données.

11.2 Identification des séries labellisées

distinction claire entre séries labellisées, séries reconnues d'intérêt général par l'ASP et autres séries (utilisation des logos transmis par l'ASP pour les séries labellisées et reconnues d'intérêt général). Possibilité d'accéder globalement aux séries labellisées (et/ou à celles reconnues par l'ASP), via un menu, une liste de ces séries... Rappel de ce qu'est la labellisation (cf. texte de présentation transmis avec les logos ASP). Liens vers les avis de l'ASP.

Dans la rubrique Données et publications du site internet est prévue une page présentant la labellisation des données, les liens vers les différentes données labellisées, ainsi que l'ensemble des avis de l'autorité de la statistique publique. Cette page pourra être référencée sur la page des Grandes tendances et sur le site de l'Opendata à l'issue de la labellisation.

11.3 Calendrier prévisionnel de publication

(si possible à au moins six mois)

Le calendrier de publication est mis à disposition des utilisateurs sur le site internet ([page dédiée](#)) en précisant pour chaque sous-produit (grandes tendances nationales, base individuelle des causes médicales au SNDS et Opendata) les dates de mises à disposition prévues pour le prochain millésime de données soit entre 18 et 21 mois après la fin de l'année des décès.

11.4 Politique de correction d'erreur

Mesures mises en place, mode d'information des utilisateurs

La politique de correction d'erreur suit le code européen des bonnes pratiques et les échanges du SSP coordonnés par l'unité qualité de l'Insee en la matière.

- Les erreurs sont corrigées dans les meilleurs délais après découverte de celles-ci, selon leur qualification
- Les corrections sont signalées dans un journal d'erreur, qui indique l'erreur corrigée, la date de correction, l'impact sur les données et l'information fournie aux utilisateurs.
- Les erreurs sont qualifiées, et selon le type d'erreur les utilisateurs sont informés.
- La mention d'erreur se fait au plus proche de la source de diffusion

Les erreurs repérées concernent à ce jour uniquement les données diffusées par le biais de l'Opendata (voir dans le bas de la page de [Note d'information](#) cette liste).

11.5 Explicitation des sigles et concepts

(concepts statistiques ou propres au domaine), le cas échéant lien vers un glossaire

La [fiche méthodologique](#) (annexe) décrit les différents indicateurs mis à disposition. Les liens sont donnés vers les différents éléments de nomenclature lorsqu'ils sont cités et notamment la shortlist européenne ainsi que les classifications internationales des maladies version 9 et 10.

11.6 Description des nomenclatures de diffusion

(avec différents niveaux d'arborescence possibles), et le cas échéant de leurs principales évolutions et de leur gestion (rétopolotatation, année en double codification...)

La seule nomenclature de diffusion ayant été modifiée sur la période est la nomenclature des régions en 2016 : l'ensemble des indicateurs est disponible dans les deux découpages géographiques depuis 1979.

La nomenclature internationale de classification des maladies utilisée pour coder l'indicateur de cause initiale de décès évolue continuellement au fil du temps pour tenir compte des nouvelles maladies (comme le COVID-19) et des besoins des utilisateurs. Depuis 1979, deux versions de la CIM ont été utilisées en France pour coder les décès :

- la CIM-9 entre 1979 et 1999 ;
- la CIM-10 depuis 2000.

La classification dans la shortlist d'Eurostat propose une table de passage entre les différentes CIM et les 86 postes de la shortlist. Mais au passage de la CIM-9 à la CIM-10, certaines maladies ont été reclassées dans des agrégats différents d'Eurostat. Par ailleurs, comme la CIM ne donne pas seulement la nomenclature de classification mais également les règles de détermination de la cause initiale : quelles maladies peuvent être retenues en cause initiale ? dans un enchaînement de causes jusqu'où remonte-t-on pour choisir la cause initiale ? La transition de la CIM 9 à la CIM 10 a conduit dans certains cas à retenir des causes initiales différentes pour un même processus morbide. Des révisions ultérieures CIM-10 ont été opérées et même si elles sont de moindre ampleur, elles ont aussi créé des ruptures de série.

11.7 Documentation articulée selon différents niveaux

(à visée pédagogique pour le grand public et méthodologique pour les spécialistes du domaine ; articulation entre information générale et information spécifique à chaque série) :

- *Liens des séries vers les informations méthodologiques et la documentation générale transverses aux séries ; précisions spécifiques à chaque série directement attachées à la série (y compris mention des sources, champ sous les tableaux ou graphiques) ;*
- *Liens des séries vers d'éventuelles publications les commentant ;*

Sur la page des grandes tendances, sont référencées les dernières publications marronnières commentant les principales évolutions de mortalité et les rapports de production.

Une partie du site, accessible via le bandeau du haut, « production des causes de décès » est dédiée à l'explication synthétique de l'objectif de la statistique sur les causes de décès, la collecte et les principaux traitements statistiques. Elle se veut accessible au grand public. Une autre rubrique du site « Documentation » est dédiée aux documents de travail et fiches méthodologiques. Une page du site internet regroupent toutes les publications scientifiques et méthodologiques du CépîDc. Enfin l'accès au site OpenData se fait via une page de présentation/ note informative.

- *Principales évolutions impactant fortement les séries (précisions attachées aux séries pour signaler des évolutions majeures dans l'objectif de limiter le risque de mauvaise interprétation de fortes évolutions, et, si disponible, lien vers une description plus complète de l'historique des évolutions – par exemple, évolutions de la législation de prestations...)*
- *Affichages de précautions d'usage éventuelles, de conseils d'utilisation*

Le site d'Opendata est accompagné d'une page présentant les principales précautions permettant un bon usage des données mises à disposition avec le rappel des ruptures dues au changement de nomenclature de classification des maladies : passage de la CIM 9 à la CIM 10 en 2000, la révision de la CIM 10 en 2019. Un travail de recensement plus détaillé des principales ruptures est en cours pour accompagner la diffusion des statistiques sur l'Opendata avec une description des principales ruptures observées.

Les changements de périmètre géographique au fil des années avec l'inclusion progressive des Droms y sont également documentés.

Sur la page Grandes tendances, les principales évolutions de méthodes de codage (rattrapage puis introduction du codage par RNP en production régulière) des dernières années (depuis 2015) sont mentionnés en note de lecture des tableaux.

11.8 Liens vers des sites partenaires

pour des éclairages complémentaires (et brève explicitation de l'articulation avec l'offre de ces partenaires)

Au-delà des liens vers les publications marronnières coécrites avec Santé publique France et la Drees répertoriées sur le site, sont également indiqués les liens vers le Data Drees qui regroupe un certain nombre de données de référence sur le sujet et le lien vers les séries longues localisées diffusées sur le site de l'Insee (chiffres-clés et séries géographiques).

11.9 Modalités d'accès aux données pour les chercheurs.

Les [modalités d'accès aux données](#) sont détaillées sur le site internet en guidant l'utilisateur suivant ses besoins et son statut :

- données en accès libre : grandes tendances et Opendata
- Pour les chercheurs, les liens vers les différents processus permettant d'accéder aux données détails (qui renvoie sur la plateforme du HDH) : SNDS (accès permanent ou non), ou via convention (avec accord Cnil préalable en suivant le processus décrit sur la plateforme du HDH pour solliciter l'avis du CESREES),
- ou demande d'extraction complémentaire non réidentifiante au CépiDc. Ces mêmes types d'information sont accessibles à partir du bandeau du haut « Demandes de renseignements » par profil d'utilisateur.

12 Gestion de la confidentialité statistique.

Secret primaire et secondaire. (se rapporte au principe n° 5 du CBPSE)

Les données collectées par le CépiDc sont régies par l'article L 2223-42 du code des collectivités territoriale indiquant que « le certificat de décès précise de manière confidentielle la ou les causes de décès ». Les actions garantissant la confidentialité des données tout au long du processus de production ont déjà été abordées dans le reste du rapport. On se concentre ici sur les séries diffusées. Les séries agrégées diffusées sur la page Grandes tendances, au niveau national ne posent pas de question. Celles représentant des agrégations plus détaillées (département de domicile et âge décennal) sur le site OpenData sont elles traitées pour garantir le secret.

En effet, outre cette obligation de confidentialité réglementaire, et même si les personnes étant décédés le RGPD ne s'applique pas, la divulgation de certaines causes de décès pourrait avoir des conséquences sur les proches de l'individu décédé par exemple en application des règles d'exclusion de garantie des assurances décès, et il s'agit de données de santé.

Certains croisements en diffusion sur l'open-data présentent un risque de réidentification d'individus notamment dans les départements avec une population moindre et sur les classes d'âge où les décès sont plus rares. Ceci serait d'autant plus facile depuis la diffusion en open data du fichier des personnes décédées de l'Insee contenant le nom et le prénom du défunt, sa date de naissance et de décès, son sexe ainsi que le lieu de domicile. C'est la raison pour laquelle une méthode d'anonymisation (floutage des données au niveau individuel par permutation) est réalisée sur l'open-data pour éviter les réidentifications des décès dont la cause initiale fait partie des causes externes. Cette méthode, respectant une diversité des causes pour un groupe donné (L-diversité) de 80% et une taille minimum du groupe (K-anonymité) de 5, satisfait les critères habituels de respect du secret pour les décès par causes externes. Pour autant, elle a deux défauts : le premier est qu'elle ne répond que partiellement à la préservation du secret statistique

pour les autres causes de décès, et le deuxième est que son choix méthodologique conduit à une sous-estimation systématique des événements rares (par exemple les décès chez les jeunes femmes).

Le CépiDc a l'intention de faire évoluer cette méthode très prochainement. Il a échangé début 2025 avec la Direction de la méthodologie et de la coordination statistique et internationale de l'Insee pour valider une nouvelle méthode d'anonymisation de la base en Opendata. Le choix s'est porté sur une méthode de *target record swapping* qui permet à la fois de cibler des observations à bruiteur car à risque d'identification et de respecter un certain nombre de marges globales déterminées par l'utilisateur. Cette méthode d'anonymisation a été développée par l'Office for National Statistics et a été utilisée dans le cadre du recensement européen de 2021 par l'Insee sur les données françaises. Les analyses des résultats de cette méthode de *swapping* sont en cours, donnant lieu à la rédaction d'une note méthodologique qui sera publiée pour informer les utilisateurs de l'Opendata. L'anonymisation sera implémentée au plus tard lors de la prochaine mise à jour de l'Opendata avec les données sur les causes des décès en 2023.

13 Moyens humains et financiers mis en œuvre pour les différentes étapes du processus statistique

recueil et codification, traitements, études et diffusion, (se rapporte aux principes n° 3 et 7 du CBPSE)

Le CépiDc comprend aujourd'hui 21 agents, et à terme 20 agents (dont 5 dont on demande le support financier à compter de 2026). L'organigramme présenté supra et détaille les principales fonctions des agents et où ils interviennent dans le processus. Le CépiDc fonctionne actuellement grâce au financement annuel versé à l'Inserm pour charge de service public (380 000 euros que l'Inserm régulièrement complète, voir dans le tableau budget ci-dessous les dotations exceptionnelles annuelles de l'ordre de 180 000 euros), à la masse salariale des ITA et agents en CDD sur support Inserm (15 postes gagés et financés) et aux financements extraordinaires versés dans le cadre du projet de rattrapage et de refonte de la statistique sur les causes de décès (2,8 millions d'euros au total).

Au-delà de la masse salariale associée aux postes, les frais annuels de fonctionnement de la structure à prévoir dès le 1er janvier 2026 sont estimés à 730 000 euros par an (voir note de demande de resoclage en annexe). Compte tenu de ce que la dotation de fonctionnement historique de l'unité était de 380 000€, ce besoin traduit la nécessité d'un rebasage de cette dotation à hauteur de 350 000 € supplémentaires par an. Ces 730 000 euros par an se décomposent en :

- 400 000 euros pour la saisie/numérisation/standardisation des certificats : recours à un prestataire, qui couvre en plus de la numérisation des certificats papier, la standardisation de certains certificats électroniques, la mise en place des changements législatifs (nouvelle version de certificat), les évolutions du portail de suivi de la saisie ;
- 70 000 euros pour les frais courants réguliers de la structure : achats d'ordinateurs portables, matériel de bureau, missions, conférences, déplacements, imprimante, petit mobilier, abonnements téléphone, poste, gratifications de stage, hébergement et évolution du site internet, formations, pour une équipe d'environ 20 personnes ;
- 65 000 euros de maintenance évolutive du système d'information du CépiDc ;
- 80 000 euros de prestation de codage du texte médical en CIM 10 externalisée
- 115 000 euros de maintien en conditions opérationnelles et l'hébergement des infrastructures du système d'information du CépiDc au sein de data centers (soit Inserm soit extérieurs)

Par ailleurs, dans le cadre du Projet de rattrapage et de refonte, un montant de 500 000 euros a été versé pour conduire une refonte fonctionnelle du système d'information (chaîne de production incomplète et obsolète). Cette refonte, indispensable, est programmée pour les deux années qui viennent mais elle n'a pas encore démarré.

La révision du budget annuel de fonctionnement demandée ici, plus l'enveloppe déjà versée pour la refonte dans le cadre du Projet de rattrapage et de refonte, permettraient d'assurer au CépiDc des conditions pérennes de fonctionnement y compris informatique.

On détaille ci-dessous les éléments de budget, ressources et dépenses des dernières années.

BUDGET

Dotations récurrentes (subvention d'Etat) des opérateurs tutelles pour chacune des 4 dernières années		
Année	Nom du financeur	Montant (k€)
2024	SE (Inserm)	380+180(dotation exceptionnelle)
2023	SE (Inserm)	380+180 (dotation exceptionnelle)
2022	SE	380
2021	SE	380

Compter aussi un financement exceptionnel pour l'aménagement des locaux Laplace débloqué en 2023 de 245keuros

Projet de rattrapage et de refonte de la statistique des causes de décès 2022-2024 – financement 2022 de 861Keuros sur COMP + complément de SCSP (MESR) spécifique étalé sur 2023-2024 de 1981 keuros (dont 500 pour la refonte informatique)		
Année	Nom du financeur	Montant (k€)
2024	MESR	620 (+200 pour refonte info)
2023	MESR	861 (+300 pour refonte info)
2022	COMP	861

Ressources propres des 4 dernières années (Grant ANR, PIA, fondation, association, INCa, IBISA, agences EU, collectivités territoriales ... et crédits exceptionnels d'établissements publics ou privés, des dons et legs) sous forme de contrat de recherche, de subventions (hors contrats de collaboration) ; cette rubrique intègre également une contribution éventuelle des UMR associées (de site le plus souvent)				
Année d'obtention	Nature et nom de la ressource	Durée du financement	Montant perçu / à percevoir (k€)	Nom du financeur
2024	SPF	12 mois	55	Santé publique France
2023	SPF	12 mois	55	Santé publique France
2022	SPF	12 mois	55	Santé publique France
2021	SPF	12 mois	50	Santé publique France
20-22	SpF	24 mois	150	DcCov

Dépenses sur les 4 dernières années : fonctionnement, maintenance et CDD		
Année	Nature des frais	Montant (k€)

2024	Fonctionnement : SE + rattrapage	376 (numérisation) + 88.5 (fonctionnement CépiDc) + 188 (Projet)
	Hébergement + Maintenance + maintenance applicative :	13,5 (région) + 17,5 (hébergement national) + 75 (maintenance applicative)
	CDD* :	436 (Projet) +10 (SPF)
2023	Fonctionnement :	475 (numérisation) + 43 (fonctionnement CépiDC) + 181 (Projet)
	Hébergement + Maintenance + maintenance applicative:	30 (accompagnement migration serveur) + assuré par DSI
	CDD* :	342 (Projet)+ 55 (SpF)
2022	Fonctionnement :	382 (numérisation) + 75 (fonctionnement CépiDc)+50 (Projet)
	Hébergement + Maintenance + maintenance applicative:	Assuré par DSI
	CDD* :	146 (Projet) +55 (SpF) + résidus DcCov
2021	Fonctionnement :	372 (numérisation) + 24(fonctionnement CépiDc)
	Maintenance :	Assuré par DSI
	CDD* :	50 (SpF) + 75 (DcCov)

*hors CDD et ITA payés sur support INSERM

14 Autres projets et objectifs des prochaines années pouvant influencer sur les séries publiées de façon indirecte

Les éléments reportés ci-dessous synthétisent les points de vigilance et les projets indiqués dans le dossier de demande de renouvellement de l'unité soumis par le CépiDc fin novembre 2024.

Pérennité des moyens - Qu'il s'agisse de moyens humains ou financiers, le CépiDc remplit sa mission actuellement grâce aux financements supplémentaires alloués dans le cadre du projet de rattrapage. Ces financements seront dépensés fin 2025 début 2026. La priorité est donc de sécuriser les moyens humains et financiers alloués à l'unité pour cette charge de service publique.

Organisation/ structuration pour sécuriser la production de la statistique sur les causes de décès en accord avec le code des bonnes pratiques en matière de statistique européenne - L'histoire récente montre que l'organisation du CépiDc, sa structuration, sa place de formation de recherche en unité de service dans l'écosystème Inserm et dans son propre écosystème mixant les acteurs de la santé, de la veille sanitaire et de la statistique publique n'avait pas permis d'éviter le défaut de production en 2018-2019 qui a conduit au lancement de la mission de rattrapage et de refonte. L'objectif est que cela ne se reproduise pas. A ce titre, la gouvernance pérenne mise en place cet automne offre un lieu pour

communiquer efficacement sur les problématiques rencontrées, avoir une compréhension partagée au sein de l'écosystème des prérequis et des critères de qualité de la statistique publique auxquels le CépiDc est soumis, et de lever les moyens d'action. Pour assurer cette sécurisation de façon durable, il faut aussi que les missions spécifiques et les critères de qualité associés soient reconnus/acceptés au sein de l'Inserm (évaluations internes de l'Inserm, campagnes de postes, relations avec la sphère informatique,...). Ces critères sont pour la plupart très proches de ceux de la recherche (indépendance, méthodologie, pertinence, objectivité, confidentialité,...). Seule la question de la pérennité/continuité (en lien avec les délais) est plus spécifique à la Statistique publique. Il faut aussi que l'Inserm assure en continu les postes gagés sur sa trajectoire d'emploi (15 aujourd'hui sur le plafond actuel, et une demande en cours de 5 supplémentaires qui augmenteront d'autant le plafond d'emploi de l'Inserm) et que les vacances de postes soient minimisées au plus de façon à assurer une production en continu.

Informatique –

Le financement du Projet de rattrapage et de refonte comprend un budget de 500 keuros dédié à une refonte du système d'information. Par refonte, on entend une refonte applicative pour revoir certains traitements de la chaîne de production obsolètes (comme la correction des certificats via interface manuelle avant intégration dans la base) pour en industrialiser d'autres (traitements automatiques de codage batch tournant actuellement sur machine locale), pour en couvrir d'autres non prévus notamment en développant des environnements dédiés (serveurs de calcul dédiés au self pour les traitements statistiques), pour sécuriser la chaîne de production (environnements tests/pré-prod...). Évidemment, des évolutions des infrastructures et de leur exploitation/supervision en découlent. Les décisions qui seront prises à la suite par l'Inserm influenceront largement sur la poursuite de la modernisation et de la pérennisation de l'activité du CépiDc. Notamment, il sera essentiel de veiller à ce que les prérequis propres à un processus de production d'une base de données statistique soient satisfaits (continuité, fluidité de service, capacités de traitement, supervision, réactivité, sécurité, intégrité....) à la cible et sur la trajectoire menant à la cible.

Il ne faut que le **périmètre d'action du service** soit élargi, changé. La mission du CépiDc est la production de la base sur les causes médicales de décès. Les expériences passées conduisant à des élargissements du champ des missions, ont au contraire nui à la mission première du service.

Fondements juridiques

- [Règlement \(CE\) n° 1338/2008 du Parlement européen et du Conseil du 16 décembre 2008 relatif aux statistiques communautaires de la santé publique et de la santé et de la sécurité au travail](#). In : *site de l'Union européenne*. [en ligne]. [Consulté le 6 novembre 2024].
- [Règlement \(UE\) n° 328/2011 de la Commission du 5 avril 2011 portant application du règlement \(CE\) n° 1338/2008 du Parlement européen et du Conseil relatif aux statistiques communautaires de la santé publique et de la santé et de la sécurité au travail, en ce qui concerne les statistiques sur les causes de décès](#). In : *site de l'Union européenne*. [en ligne]. [Consulté le 6 novembre 2024].
- [Article L2223-42 du Code général des collectivités locales](#). In : *site de Légifrance*. [en ligne].
- [Article R2213-1 du Code général des collectivités locales](#). In : *site de Légifrance*. [en ligne].
- [Loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques](#). In : *site de Légifrance*. [en ligne]. Mis à jour le 25 mars 2019.

- [Arrêté du 29 mai 2024 relatif aux deux modèles du certificat de décès](#). In : *site de Légifrance*. [en ligne].

Références

Documents de travail et notes du CépiDc

Boulet P, Hebbache Z, Robert A, Zambetta E, Razakamanana D, Aubineau Y, et al. [Rapport de Production sur les causes de décès en 2022](https://www.cephdc.inserm.fr/sites/default/files/2025-03/DT_CEPIDC_N5_Rapport%20de%20production%202022.pdf) [Internet]. CépiDC - INSERM; 2025 p. 100. Report No.: 5. Available from: https://www.cephdc.inserm.fr/sites/default/files/2025-03/DT_CEPIDC_N5_Rapport%20de%20production%202022.pdf

Godet F, Définitions des indicateurs sur les causes de décès utilisés sur le site du CépiDc. Fiche méthodologique. Janvier 2025 https://www.cephdc.inserm.fr/sites/default/files/2025-01/FM_Indicateurs.pdf

Hebbache Z, Boulet P, Robert A, Zambetta E, Razakamanana N, Coudin E, et al. [Rapport de production: année de décès 2021](https://www.cephdc.inserm.fr/documentation/rapport-de-production-annee-de-deces-2021) [Internet]. 2024. (Document de travail du CépiDc). Report No.: 4. Available from: <https://www.cephdc.inserm.fr/documentation/rapport-de-production-annee-de-deces-2021>

Zambetta E, Razakamanana N, Robert A, Clanché F, Rivera C, Martin D, et al. Codage des causes de décès de 2018 et 2019 en CIM10 - Approche combinant deep learning, système expert et codage manuel ciblé [Internet]. Centre d'épidémiologie sur les causes médicales de décès; 2023 Sep. (Document de travail du CépiDc). Report No.: 2 (version française). Available from: <https://www.cephdc.inserm.fr/documentation/codage-des-causes-de-deces-de-2018-et-2019-en-cim10-approche-combinant-deep-learning-systeme-expert-et-codage-manuel-cible-document-de-travail-cepidc-n22023>

Zambetta E, Razakamanana N, Robert A, Clanché F, Martin D, Hebbache Z, et al. Combining a deep-learning-based approach, rule-based automated expert system and targeted manual coding for ICD-10 cause of death coding of French death certificates in 2018 - 2019 [Internet]. Centre d'épidémiologie sur les causes de décès - CépiDc- Inserm; 2023 Sep. (Document de travail du CépiDc). Report No.: 2/2023. Available from: <https://www.cephdc.inserm.fr/documentation/codage-des-causes-de-deces-de-2018-et-2019-en-cim10-approche-combinant-deep-learning-systeme-expert-et-codage-manuel-cible-document-de-travail-cepidc-n22023>

Autres publications méthodologiques

Clanché, François, Razakamanana, Nirintsoa, Coudin, Élise et Robert, Aude, 2023. [Les statistiques provisoires sur les causes de décès en 2018 et 2019 - Une nouvelle méthode de codage faisant appel à l'intelligence artificielle](#). In : *Drees Méthodes*. [en ligne]. Mars 2023. N° 8.

Metatdata Causes of Death : France
https://ec.europa.eu/eurostat/cache/metadata/EN/hlth_cdeath_simscd_fr.htm

Zambetta, Elisa, Nirintsoa Razakamanana, Aude Robert, François Clanché, Cecilia Rivera, Diane Martin, Zina Hebbache, Rémi Flicoteaux and Elise Coudin. "Combining deep neural networks, a rule-based expert system and targeted manual coding for ICD-10 coding causes of death of French death certificates from 2018 to 2019." *International journal of medical informatics* 188 (2024): 105462 .
<https://www.sciencedirect.com/science/article/abs/pii/S1386505624001254?via%3Dihub>

Analyses marronnières

Cadillac, Manon, Fouillet, Anne, Rivera C, Clanché F, Coudin E. Grandes causes de décès en France en 2021 : une année encore fortement marquée par le Covid-19. *Etudes et Résultats*. 2023 Dec;1288

Cadillac, Manon, Fouillet, Anne, Rivera, Cecilia, Coudin, Elise. Les causes de décès en France en 2022 : recul du Covid-19 et hausse des maladies respiratoires. *Etudes et résultats*. 2024 Oct;1312.

Fouillet, Anne, Cadillac, Manon, Rivera, Cecilia, Coudin, Elise. Grandes causes de mortalité en France en 2022 et tendances récentes. *Bull Epidémiol Heb [Internet]*. 2024 Oct;18:388–404. Available from: http://beh.santepubliquefrance.fr/beh/2024/18/2024_18_html

Fouillet A, Ghosn W, Rivera C, Clanché F, Coudin E. Grandes causes de décès en 2021 et tendances récentes. *Bulletin épidémiologique hebdomadaire*. 2023 Dec;(26):554–69.

Fouillet A, Ghosn W, Naouri D, Coudin E. Covid-19 : troisième cause de décès en France en 2020, quand les autres grandes causes baissent. *Bull Epidémiol Hebd*. 2022;Cov_16:2–15.

Fouillet A, Ghosn W, Naouri D, Coudin E. Covid-19 : troisième cause de décès en France en 2020, quand les autres grandes causes de décès baissent. *Etudes et résultats*. 2022;(1250):1–7.

Références citées dans le rapport

Aouba, Péquignot, Camelin, Jouglu, Evaluation de la qualité et amélioration de la connaissance des données de mortalité par suicide en France métropolitaine, 2006. *Bull Epidémiol Hebd*. 2011, n°. 47-48, p. 497-500

Clanché, François, Razakamanana, Nirintsoa, Coudin, Elise, Robert, Aude. Les statistiques provisoires sur les causes de décès en 2018 et 2019 - Une nouvelle méthode de codage faisant appel à l'intelligence artificielle. *Drees - Méthodes*. 2023;8.

Coudin E, Fouillet, A. De l'importance du volet médical du certificat de décès pour la connaissance des causes de décès en France,. *La Presse Médicale Formation [Internet]*. 2025;6. Available from: sciedirect.com/science/article/pii/S2666479825000084

Coudin, E, Robert, A. Les statistiques sur les causes de décès : classer et coder ... dans la classification internationale des maladies. *Courrier des statistiques N12*, 2024, décembre. <https://www.insee.fr/fr/information/8264550?sommaire=8264562>

Falissard L, Morgand C, Roussel S, Imbaud C, Ghosn W, Bounebache K, et al. A Deep Artificial Neural Network-Based Model for Prediction of Underlying Cause of Death From Death Certificates: Algorithm Development and Validation. *JMIR Med Inform*. 2020 Apr 28;8(4):e17125.

Iris Institute, 2024. [Iris software](#). In : *site de Federal Institute for Drugs and Medical Devices*. [en ligne].

Lamarche-Vadel, A., Pavillon, G., Aouba, A. et al. Automated comparison of last hospital main diagnosis and underlying cause of death ICD10 codes, France, 2008–2009. *BMC Med Inform Decis Mak* **14**, 44 (2014). <https://doi.org/10.1186/1472-6947-14-44>

Organisation mondiale de la santé, 2008. [Classification statistique internationale des maladies et des problèmes de santé connexes, dixième révision, Volume 2](#). In : *site de l'Organisation mondiale de la santé*. [en ligne]. Édition 2008.

Organisation mondiale de la santé, 2008. [Classification statistique internationale des maladies et des problèmes de santé connexes, dixième révision, Volume 2](#). In : *site de l'Organisation mondiale de la santé*. [en ligne]. Édition 2019. https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf

Pavillon, Gérard et Laurent, Françoise, 2003. [Certification et codification des causes médicales de décès](#). In : *Bulletin Épidémiologique Hebdomadaire*. [en ligne]. 8 juillet 2003, mis à jour le 30 août 2019, n° 30-31, p. 134-8. [Consulté le 6 novembre 2024].

Pierre, O. (2023). Évaluation de l'effet du nouveau modèle de certificat de décès sur la statistique nationale des décès par suicide. Thèse pour le diplôme d'état de docteur en médecine, université Paris-Cité

Rey, Grégoire, 2016. [Les données des certificats de décès en France : processus de production et principaux types d'analyse](#). In : *La Revue de Médecine Interne*. [en ligne]. Octobre 2016. Volume 37, n° 10, pp. 685-693. [Consulté le 6 novembre 2024].

Rey, G., Aouba, A., Pavillon, G. et al. Cause-specific mortality time series analysis: a general method to detect and correct for abrupt data production changes. *Popul Health Metrics* 9, 52 (2011). <https://doi.org/10.1186/1478-7954-9-52>

Et ici la liste des publications scientifiques auxquelles des membres du CépiDc ont participé au titre de leur activité au CépiDc : <https://www.cepiddc.inserm.fr/publications-scientifiques>

Liste des annexes

- A1-2024-036_ASP_M Elbaum_Dossier soumission comité label statistique publique- Lettre de demande de soumission du rapport
- A2-reglement CE 1338/2008 du parlement européen et du conseil relatif aux statistiques communautaires de la santé publique et de la santé et de la sécurité au travail
- A3-reglement UE 328/2011 de la commission portant application du règlement CE 1338/2008 (...) en ce qui concerne les causes de décès
- A4- Rgl't intérieur CépiDc-2024 VF signee.pdfLP : règlement intérieur du CépiDc
- A5 – Note préparatrice du comité stratégique du 7 octobre 2024
- A5b – CR du comité stratégique du 7 octobre 2024
- A6 – Note préparatrice et CR du comité stratégique du 4 février 2025
- A7- lettre du dg de l'Insee au pdg de l'Inserm sur le statut d'ONA du CépiDc
- A8- 2024 European Manual for Causes of death Statistics
- A9 – Fiche méthodologique sur les indicateurs
- A10- Lettre du PDG de l'Inserm témoignant à l'ASP du processus de recrutement de l'actuelle directrice du CépiDc
- A10b Fiches de poste de Directeur/ Directrice du CépiDc 2022
- A10c Fiches de poste de Directeur/ Directrice du CépiDc 2025
- A11 Coudin- Fouillet La Presse Medicale Formation - De l'importance du volet médical du certificat de décès pour la connaissance des causes de décès en France
- A12 Rapport de production des causes de décès en 2022
- A13 et A 13b Certificats de décès en vigueur (volet administratif et volet médical)
- A14 et A14b SFD et CCTP de l'application CertDc (+ tableur recensant les règles de gestion de l'application : filtres, etc.)
- A15 Schéma Insee circuit d'un décès
- A16 CCTP du marché de saisie numérisation standardisation 2023
- A17 Note proposition automatisation et intégration a l'intégration
- A18 Règles saisies Causes détaillées appliquées par Luminess

-A19 - Lettre D-24-015060 - Comité stratégique de la statistique sur les causes de décès

-A20 - note déroulé calendrier BEH et ER 2022 pour publication simultanée

-A21 – note de demande de resoclage 2026 – Drees DGS DGRI Inserm

-A22 – Lettre signée du pdg de l’Inserm s’engageant sur le calendrier de diffusion

Liste des annexes sous forme de tableaux

- T1 - Tableau des séries pour lesquelles est demandée la labellisation
- T2 - Contrôles et filtres à la collecte spécifiés dans l’application CertDc
- T3 - Contenu de la table principale mise au SNDS pour l’année 2022 : variables, valeurs manquantes, répartition des modalités prises
- T4 - Autoévaluation sur les critères du Code des pratiques
- T5 - Etat des lieux des ruptures de séries.