Centre d'épidémiologie sur les causes médicales de décès – Institut national de la santé et de la recherche médicale – Documents de travail du CépiDc

Rapport de production

sur les causes médicales de décès survenus en 2023

Pierre Boulet, Zina Hebbache, Aude Robert, Elisa Zambetta, Daniel Razakamanana, Yann Aubineau, Fanny Godet
CépiDc-Inserm

Décembre 2025

Document de travail du CépiDc N°06

Ces documents de travail ne reflètent pas la position de l'Inserm et n'engagent que leurs auteurs.

Résumé

Ce rapport de production revient sur l'ensemble des traitements réalisés par le CépiDc pour produire la base de données sur les causes médicales de décès survenus en 2023. Le codage en CIM-10 des causes de décès survenus en 2023 en France combine un codage automatique par le système expert de codage IRIS/MUSE, des prédictions par des algorithmes d'apprentissage profond, et un codage manuel en partie ciblé sur les certificats d'intérêt spécifique pour la recherche et la santé publique. Une attention particulière porte sur l'évaluation de la cohérence entre le codage obtenu lors de cette campagne à trois modes de codage et celui qui aurait été obtenu dans une campagne traditionnelle qui ne mobiliserait pas les réseaux de neurones profonds. Les spécificités de la base de données individuelles accessibles via le SNDS sont aussi précisées.

Mots-clés : causes de décès, mortalité, CIM-10, 2023, production statistique des données sur les causes de décès

Abstract

This production report details the treatments carried out by CépiDc to produce the database on medical causes of death in 2023. The coding of causes of death in ICD-10 in 2023 in France combines automatic coding by the IRIS/MUSE expert coding system, predictions by *deep learning* algorithms, and manual coding targeted in particular at certificates of specific interest for research and public health. Particular attention is paid to assessing the coherence between this 3 coding-mode campaign and a traditional campaign that would not mobilize deep neural networks. The database of individual data accessible via the SNDS are also presented.

Keywords: causes of death, mortality, ICD10, 2023, statistical production of CoD data

Table des matières

1 INTRODUCTION 5

2 COLLECTE 5

2.1 EXHAUSTIVITE 5
2.2 SYNCHRONISATION AVEC LES DONNEES DE L'INSEE 6
2.2.1 Répartition des responsabilités institutionnelles 6
2.2.2 Principe de la synchronisation 6
2.2.3 Limites et difficultés du dispositif 6
2.2.4 Résultats de la synchronisation 7
2.3 ORIGINE DES CERTIFICATS : ELECTRONIQUE OU PAPIER 8
2.4 VERSION DE CERTIFICATS : 2017 OU 1997 9
2.5 TYPE DE CERTIFICAT : GENERAL OU NEONATAL 9
2.6 VOLET MEDICAL COMPLEMENTAIRE (VMC) 9

3 CODAGE 10

3.1 UNE METHODE COMBINANT TROIS MODES DE CODAGE 10
3.2 REPARTITION DES MODES DE CODAGE 11
3.3 ALLOCATION DES CERTIFICATS VERS LA REPRISE MANUELLE OU L'INFERENCE PAR RNP 12
3.4 CODAGE PAR SYSTEME EXPERT IRIS/MUSE 13
3.5 CODAGE ET REPRISE PAR L'EQUIPE DE CODAGE 14
3.6 CODAGE IA (DEEP LEARNING) 16
3.6.1 Réseau de neurones transformer pour coder la séquence de causes 17
3.6.2 Un sur-modèle pour retenir une cause initiale parmi les différentes propositions possibles 19
3.7 VERIFICATIONS 21
3.7.1 Vérification de cohérence 22
3.7.2 Décès spécifiques 22
3.7.3 Difficultés de codage connues liées au logiciel Iris 24
3.7.4 Vérification de la bonne application des nouvelles règles de 2016 25
3.7.5 Vérifications liées aux évolutions du dictionnaire 25
3.7.6 Vérifications pour raisons multiples 27

4 EVALUATION DE LA STRATEGIE DE CODAGE 28

4.1 POPULATION TEST DE REFERENCE 28
4.2 ACCURACY, COHERENCE GLOBALE ET COMPARAISON A LA CAMPAGNE PRECEDENTE 29
4.3 PRECISION, RAPPEL ET ECARTS D'EFFECTIFS 31
4.4 APPORT DES ETAPES DE REPRISE CIBLEE SUR LA PERFORMANCE GLOBALE 36

5 EVOLUTIONS DE CODAGE 37

5.1 NOUVEAUTES RELATIVES AUX RECOMMANDATIONS DE L'OMS 37
5.2 NOUVEAUTES INTERNES AU CEPIDC 37
5.3 MISES A JOUR DU DICTIONNAIRE DES EXPRESSIONS NOSOLOGIQUES UTILISE PAR IRIS MUSE 38

6 SPECIFICITES DE LA BASE DE DONNEES INDIVIDUELLES 38

6.1 VARIABLE SUR LE MODE DE CODAGE UTILISE POUR CODER CE CERTIFICAT 38

7 REFERENCES 38

8 ANNEXE 40

8.1 VERSION 2017 DU VOLET MEDICAL DU CERTIFICAT DE DECES GENERAL (28 JOURS ET PLUS) 40
8.2 VERSION 2017 DU VOLET MEDICAL DU CERTIFICAT DE DECES NEONATAL (MOINS DE 28 JOURS) 41
8.3 CALENDRIER DE PRODUCTION DE L'ANNEE DE DECES 2023 42
8.4 METHODE DE CIBLAGE DES CERTIFICATS A CODER MANUELLEMENT SUR LA BASE DE PREDICTION IA 44
8.4.1 Ciblage des certificats sur la qualité des prédictions de la cause initiale 44
8.4.2 Ciblage des certificats sur la qualité des prédictions des causes associées 47
8.5 DESCRIPTION DES BASES D'ENTRAINEMENT ET DES BASES DE TEST DES MODELES 51
8.6 DEFINITIONS DETAILLEES DES VERIFICATIONS ET EVOLUTIONS PAR RAPPORT A 2022 52
8.6.1 Evolution des vérifications 52
8.6.2 Décès entre 28 jours et 15 ans identifiés dans les décès sensibles ou à vérifier 57
8.6.3 Certificats avec mention de Covid-19 vérifiés 59
8.7 PRECISIONS SUR LES MODELES TRANSFORMERS 59
8.7.1 Entrainement / validation / test 59
8.7.2 Description des modèles 60
8.7.3 Programmes du Transformers 61
8.8 PRECISIONS SUR LE MODELE QUI SELECTIONNE LA CAUSE INITIALE ENTRE DIFFERENTES PROPOSITIONS 64
8.8.1 Base d'apprentissage 65
8.8.2 Problème de classification / variable cible 65
8.8.3 Modèle BiLSTM 66
8.8.4 Résultats et analyse de performance 69
8.9 DICTIONNAIRE DE VARIABLES DANS LE SNDS 71
8.9.1 Table Ident 71
8.9.2 Table des causes 78
8.10 EVALUATION DE LA COHERENCE DES CAUSES MULTIPLES AVEC UNE CAMPAGNE DE CODAGE A TROIS MODES 79
8.10.1 Score BLEU 79
8.10.2 Population test de référence 83
8.10.3 Performance sur la séquence de cause multiple 84
8.11 CONSTITUTION DE LA POPULATION TEST DE REFERENCE 89
8.11.1 Quelles données sont disponibles pour la campagne 2023 ? 89
8.11.2 Comment ont-été tirés les échantillons de la campagne 2023 ? 90
8.11.3 Compléter la population test de référence 91
8.12 LISTE DES TABLEAUX 93
8.13 LISTE DES FIGURES 94

1 Introduction

Ce rapport présente les spécificités du traitement des certificats des décès survenus en 2023. Il complète ainsi le document « Statistiques sur les causes de décès de A à Z » plus général sur le processus de production de ces données et fait suite aux rapports de production précédents sur les décès des années 2022, 2021 et 2018/2019 [Hebbache *et al.*, 2025, 2024 et 2023].

La base de données pour les décès de 2023 a été produite dans un contexte de stabilisation de l'articulation de la production autours de trois modes de codage : automatique, deep learning et manuel. Ce rapport donne des indicateurs liés à la collecte des données et décrit le processus de codage mobilisé en 2023. Les données sur les causes de décès survenus en 2023 ont été diffusées en juin 2025 à Eurostat, puis à l'OMS. La base de données individuelles a été mise à disposition dans le système national des données de santé (SNDS) fin juillet 2025 : elle est également décrite dans le présent document. Ces diffusions ont été accompagnées de deux analyses des grandes causes de décès en 2023 [Godet *et al.*, 2024 ; Fouillet *et al.* 2024].

2 Collecte

2.1 Exhaustivité

En 2023, 639 547 décès survenus sur le territoire français ont été comptabilisés par l'Insee[1]. Pour la même année, le CépiDc a reçu 639 480 volets médicaux et volets médicaux complémentaires et en a conservé 626 697 ; soit 98 % (Tableau 1). Le taux de collecte s'élève donc à 97,99 % en 2023, soit un niveau très légèrement supérieur à celui des années précédentes (97,93 % en 2022, 97,96 % en 2021 et 97,79 % en 2020).

Tableau 1. Exhaustivité de la collecte des volets médicaux (VM) pour les décès survenus entre 2021 et 2023

	2023	**2022**	**2021**
Nombre de décès comptabilisés par l'Insee	639 547	675 513	662 149
Nombre de volets médicaux reçus par le CépiDc	639 480	670 099	659 381
Nombre de volets médicaux supprimés	12 783	8 541	10 681
Nombre de volets médicaux non reçus	12 850	13 955*	13 449
Part de volets médicaux non reçus (en %)	2,01 %	2,07 %	2,03 %

Note de lecture : en 2023, l'Insee a comptabilisé 639 547 décès et le CépiDc a reçu 639 480 volets médicaux. Une fois 12 783 certificats correspondant à des doublons supprimés, il manquait 12 850 volets médicaux, ce qui représente 2 % des décès enregistrés par l'Insee.
** : le chiffre présenté ici diffère de celui présent dans le rapport de production sur les données 2022 et corrige une erreur détectée post publication.*
Champ : ensemble des décès survenus sur le territoire français, en 2021, 2022 et 2023
Sources : Envoi Insee annuel stabilisé ; CépiDc, bases des décès survenus en 2021, 2022 et 2023

[1] C'est le nombre de décès dénombrés dans les fichiers fournis par l'Insee, en entrée de la phase de synchronisation. Il intègre notamment des décès réceptionnés par l'Insee après qu'il a validé ses données (et donc non pris en compte dans les publications).

2.2 Synchronisation avec les données de l'Insee

En France, la gestion des décès repose sur deux documents complémentaires relevant du certificat de décès. Ils sont administrés par deux institutions distinctes de manière à garantir la confidentialité des données médicales individuelles.

2.2.1 Répartition des responsabilités institutionnelles

L'Inserm est chargé de la collecte et du traitement des volets médicaux des certificats de décès, obligatoirement remplis par un médecin ou une infirmière lors de la constatation du décès. Le volet médical, non nominatif, est adressé à l'Inserm directement en cas de certification dématérialisée ou *via* les ARS pour les certificats papier. L'Institut reçoit également le bulletin de décès (dit bulletin B7) pour les certificats papier, transmis de manière anonyme. Ces informations servent à établir la statistique nationale des causes médicales de décès.

L'Insee assure la gestion du Répertoire national d'identification des personnes physiques (RNIPP), reflet des registres d'état civil tenus par les communes. Ce répertoire est mis à jour quotidiennement à partir des bulletins statistiques de décès (B7bis) adressés par les mairies, permettant d'établir la statistique nationale démographique.

Ce double circuit assure la protection des données personnelles tout en maintenant la qualité statistique des deux volets, démographique (Insee) et médical (Inserm).

2.2.2 Principe de la synchronisation

La synchronisation Insee–Inserm consiste à rapprocher les décès enregistrés par chacun des deux organismes, afin d'assurer la cohérence des comptages transmis à Eurostat.

Cette comparaison permet de :

- garantir que le nombre total de décès enregistrés par les deux instituts reste cohérent au niveau national ;
- vérifier la complétude des données de chaque institut grâce à celles de l'autre ;
- identifier et corriger les divergences éventuelles (décès connus de l'un mais pas de l'autre).

2.2.3 Limites et difficultés du dispositif

Si ce système garantit la confidentialité des informations médicales et la fiabilité du contenu des certificats, il présente aussi certaines limites :

- les deux institutions ne recensent jamais exactement le même nombre de décès : l'Insee peut enregistrer des décès sans certificat médical correspondant, tandis que l'Inserm peut recevoir des certificats que les mairies n'ont pas transmis à l'Insee, ou ne pas recevoir certains certificats oubliés par une ARS ;
- le nécessaire rapprochement entre les bases de données est complexe : le volet médical est anonyme, et les variables utilisées pour l'appariement (date et lieu de naissance, date et lieu de décès, sexe, âge, etc.) peuvent différer entre les sources, car elles proviennent de saisies distinctes (médecins pour l'Inserm, officiers d'état civil pour l'Insee et l'Inserm pour les bulletins B7 et B7 bis).

2.2.4 Résultats de la synchronisation

Les 12 783 volets médicaux supprimés suite à la synchronisation correspondent majoritairement à des doublons, c'est-à-dire à des cas où le décès a fait l'objet de plusieurs volets médicaux. Il y a aussi quelques cas où des certificats n'ont pas été appariés à un décès comptabilisé par l'Insee : il peut s'agir de volets médicaux pour lesquels les données individuelles permettant l'appariement étaient incomplètes ou erronées.

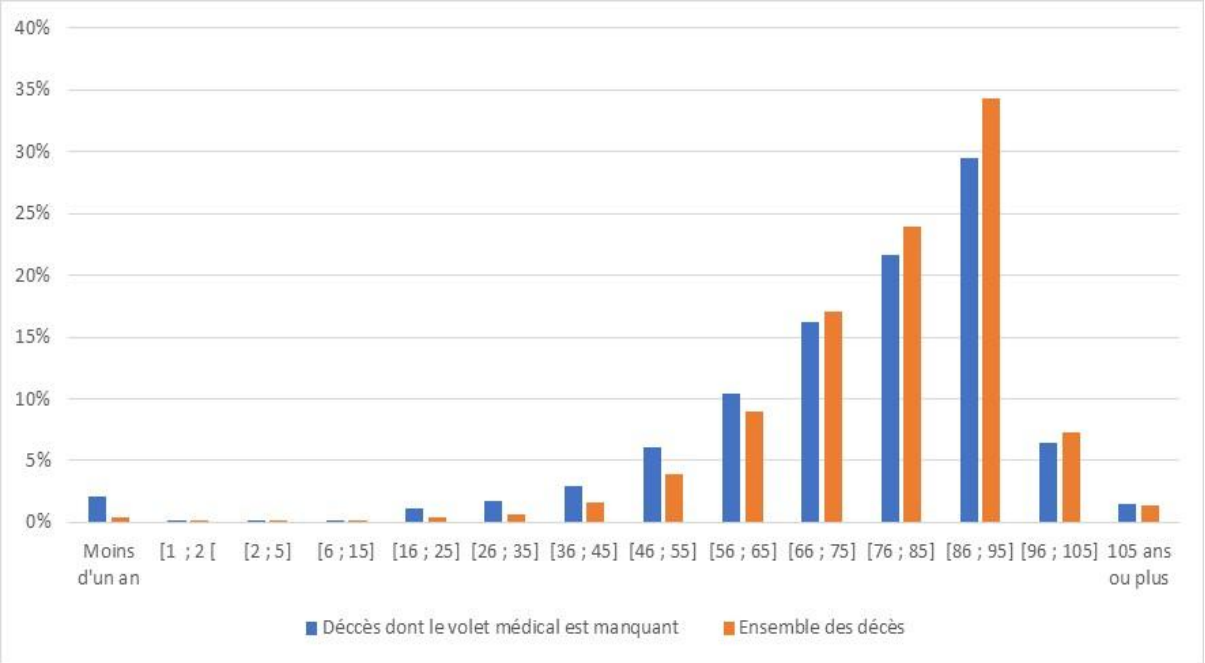
Tableau 2. Exhaustivité de la collecte des volets médicaux (VM) pour les décès survenus entre 2021 et 2023

	2023	**2022**	**2021**
Nombre de décès comptabilisés par l'Insee	639 547	675 513	662 149
Nombre de volets médicaux reçus par le CépiDc	639 480	670 099	659 381
Nombre de volets médicaux supprimés	12 783	8 541	10 681
Nombre de volets médicaux non reçus	12 850	13 955*	13 449
Part de volets médicaux non reçus (en %)	2,01 %	2,07 %	2,03 %

Note de lecture : en 2023, l'Insee a comptabilisé 639 547 décès et le CépiDc a reçu 639 480 volets médicaux. Une fois 12 783 certificats correspondant à des doublons supprimés, il manquait 12 850 volets médicaux, ce qui représente 2 % des décès enregistrés par l'Insee.
** : le chiffre présenté ici diffère de celui présent dans le rapport de production sur les données 2022 et corrige une erreur détectée post publication.*
Champ : ensemble des décès survenus sur le territoire français, en 2021, 2022 et 2023
Sources : Envoi Insee annuel stabilisé ; CépiDc, bases des décès survenus en 2021, 2022 et 2023

Pour 2 % des décès, aucun volet médical n'a été reçu, et donc aucune cause n'a été déclarée. Ces 12 850 décès sont présents dans la base de données du CépiDc avec une cause initiale de décès codée à R99 (cause indéterminée) et avec la mention « pas de certificat » dans le texte des causes. Il s'agit dans 54,3 % des cas d'hommes alors qu'ils représentent 49,8 % de l'ensemble des décès de 2023. Il s'agit aussi de défunts généralement plus jeunes que dans l'ensemble des décès (Figure 1).

Figure 1. Répartition de l'ensemble des décès et des décès manquants par groupe d'âge en 2023



Note de lecture : en 2023, 2,1 % des certificats non reçus suite à la synchronisation étaient des décès de personnes de moins d'un an alors que les décès de cette tranche d'âge représentent 0,4 % de l'ensemble des décès pour cette année.
Champ : ensemble des décès survenus sur le territoire français, en 2023
Sources : CépiDc, bases des décès survenus en 2023

Tous les décès dont la cause initiale est indéterminée (code R99) ne sont pas forcément des certificats non reçus au CépiDc. Il peut simplement s'agir de certificats dont la partie cause de décès n'a pas été complétée par le médecin ou a été insuffisamment renseignée pour identifier une cause initiale de décès informative. En 2023, 42 % des causes initiales indéterminées étaient des certificats non reçus, cette part est stable par rapport à 2022.

2.3 Origine des certificats : électronique ou papier

Pour l'année 2023, sur l'ensemble des volets médicaux reçus et conservés au CépiDc, 363 928 (58 %) étaient au format papier et 262 769 (42 %) étaient d'origine électronique. La part de certificats électroniques est en nette hausse par rapport à 2022 (36 %), même si le niveau atteint n'est pas encore satisfaisant.

L'année 2023 marque aussi le changement de prestataire pour la collecte des certificats électroniques

Les causes rédigées par le médecin sur les certificats papiers sont saisies par un prestataire qui applique des règles de saisie précises permettant de faciliter ensuite leur codage automatique (corrections de fautes d'orthographe, suppression des articles, etc.). A partir des décès 2018, ces règles de standardisation sont également appliquées sur les causes des certificats électroniques rejetés par le système expert de codage. Dans ce cas, le prestataire de saisie applique les mêmes règles de standardisation que celles appliquées sur les formulaires papier.

2.4 Version de certificats : 2017 ou 1997

En 2016, lors de la parution du volume 2 de la classification internationale des maladies (CIM-10), qui est le volume comprenant toutes les nouvelles règles de codage, l'organisation mondiale de la santé (OMS) a proposé un nouveau modèle de certificat, que la France a mis en place en 2018.

L'arrêté du 17 juillet 2017[2] (Annexe 8.1) institue ainsi deux nouveaux modèles de certificat (certificat dit « certificat 2017 ») à compter du 1er janvier 2018. Le certificat pour les décès intervenus à partie du 28ème jour (certificat dit « certificat 2017 ») contient notamment des informations complémentaires concernant les circonstances apparentes de décès : le certificateur peut ainsi renseigner s'il semble s'agir d'un suicide, d'un accident, d'une atteinte à la vie d'autrui ou d'une mort naturelle (maladie) en cochant une case. Dans le certificat néonatal (Annexe 8.2), le certificateur peut renseigner l'enchaînement causal ayant conduit au décès.

En papier, il persiste dans les faits une cohabitation des deux versions (1997/2017) pour des raisons pratiques et logistiques, les médecins utilisant les stocks de certificats ancien modèle encore en leur possession pendant plusieurs années. Sur l'ensemble des volets médicaux reçus et conservés au CépiDc pour l'année 2023, 590 353 (94 %) correspondent à la version de 2017 et 36 344 (6 %) à la version 1997, contre respectivement 608 868 (92 %) et 52 690 (8 %) pour 2022. Si l'on se concentre sur les certificats papiers, la version 1997 représente 10 % de ces certificats reçus et conservés en 2023, contre 12 % en 2022.

2.5 Type de certificat : général ou néonatal

En 2023, sur l'ensemble des volets médicaux reçus et conservés au CépiDc, 2 061 étaient des volets médicaux néonatals, dont 56 % rédigés électroniquement. Par ailleurs, 83 de ces volets médicaux ont été utilisés pour des décès de plus de 28 jours. Inversement, 8 décès chez des bébés de moins de 28 jours ont été certifiés en utilisant un modèle général.

2.6 Volet médical complémentaire (VMC)

Un volet médical complémentaire (VMC) doit être rédigé en cas d'investigations médicales ou médico-légales *post-mortem*. Il est identique au volet médical initial et est complété sur la plateforme de certification électronique. Il se substitue au volet médical initial s'il apporte davantage d'informations concernant les causes du décès. Les VMC sont collectés depuis du 1er janvier 2018.

En 2023, le CépiDc a pu prendre en compte 2 310 VMC, c'est 861 VMC de moins qu'en 2022. Près de 80 % des certificats reçus dans ce cadre proviennent de l'institut médico-légal (IML) de Paris (1 843 certificats en 2023, contre 2 573 en 2022), qui transmet l'ensemble des données issues de son système d'information directement au CépiDc : les données ainsi transmises n'ont pas le même format que le volet médical habituel car l'IML renseigne directement le code CIM-10 de la cause du décès telle que le médecin légiste la définit. Les 467 autres VMC pris en compte ont été collectés *via* l'application de certification électronique ; leur nombre est en baisse par rapport 2022 (598).

Il existe donc une forte disparité géographique sur la collecte des VMC car leur complétion est faible dans les IML de province.

[2] https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000035388290

3 Codage

Le codage des données de l'année 2023 s'est déroulé de juin 2024 à mars 2025 en parallèle de la finalisation du codage de l'année 2022. L'annexe 8.3 détaille l'ensemble du calendrier de production 2023.

Depuis le traitement des données de 2018, le processus de codage combine trois modes de codage : le codage automatique par le système expert IRIS/MUSE, le codage manuel en interactif de certificats ciblés et le codage automatique intégrant des algorithmes de *deep learning* capables de prédire les causes associées et la cause initiale de décès [Zambetta *et al.*, 2023 ; Coudin et Robert, 2024 ; rapports de production précédents]. Cette stratégie de codage mise en place par le CépiDc se différencie des campagnes de codage dite « traditionnelles » à l'œuvre précédemment (jusqu'aux données du millésime 2020[3]) et correspondant à du codage automatique par le système expert et pour le reste, à du codage manuel uniquement.

Dans cette section, nous décrivons comment les modes de codage interagissent et donnons les spécificités associées à chaque mode de codage pour l'année 2023. Nous décrivons enfin l'étape de « vérification » qui permet de valider ou de corriger le codage de certaines données ciblées[4] - et indiquons comment elle a été mise en place pour l'année 2023.

Le codage automatique par IA et l'évaluation globale du système mixte sont présentés dans la suite.

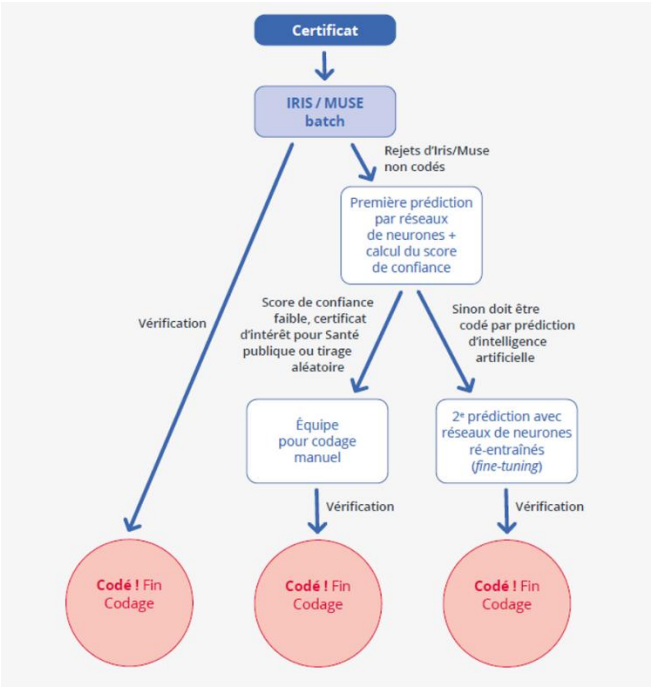
3.1 Une méthode combinant trois modes de codage

La stratégie de codage mise en place par le CépiDc combine trois modes de codage (Figure 2) [Coudin et Robert, 2024]. L'ensemble des certificats de décès est traité par le système expert Iris/Muse pour une première tentative de codage automatique. En cas de rejet, le codage manuel va se concentrer sur une partie des certificats dont le volume est fixé en fonction des ressources disponibles pour assurer la production dans les délais. La reprise manuelle concerne des tirages aléatoires, des décès sensibles et des catégories de causes de décès pour lesquelles le troisième mode de codage (IA ou inférence par RNP) est jugé comme étant moins performant. Les autres certificats sont codés par inférence d'un réseau de neurones profond (RNP). L'inférence par RNP peut être combinée avec le système expert IRIS/MUSE pour la détermination de la cause initiale. Les modèles de *deep learning* sont donc utilisés à deux moments du traitement des certificats :

- le ciblage des certificats pour le codage manuel (on utilise alors les modèles mis en œuvre lors de la production annuelle précédente) : les certificats sont ciblés selon les catégories les moins bien prédites par l'IA pour la cause initiale et par ordre de priorité selon leur score de confiance.
- Puis la prédiction du codage des causes de décès pour les certificats restant à coder après la phase de codage manuel, réalisée en toute fin de production afin d'enrichir les modèles avec le codage manuel de l'année en cours de codage.

[3] Les données 2020 ont été produites avant les données des millésimes 2018 et 2019, du fait de la crise sanitaire.
[4] Notamment les données faisant intervenir de nouvelles règles de codage, ou celles pour lesquelles des imperfections du codage automatique par système expert ont été relevées, ou encore celles concernant des décès à fort enjeu de santé publique.

Figure 2. Stratégie de combinaison des trois modes de codage



En 2023, les modèles ayant été utilisés pour cibler des certificats sont ceux entrainés en mai 2024, aussi utilisés pour les prédictions finales des données 2022. Pour la prédiction des causes de décès des certificats restant à coder après codage manuel et automatique, les modèles ont été ré-entraînés en mars 2025 en prenant en compte les certificats codés jusqu'alors.

L'étape de « vérification » consiste à valider ou corriger certaines données dont le codage semble *a priori* improbable et identifiées selon des règles déterministes préétablies (par exemple, « maladie d'Alzheimer » sur le certificat d'une personne jeune). Ces vérifications portent sur certaines catégories de certificats et sont réalisées par l'équipe de codage. Elles peuvent porter sur du codage automatique Iris/Muse, du codage manuel ou encore sur le codage IA. Pour le codage IA, ces vérifications sont en pratique du codage manuel : pour des raisons technico-fonctionnelles, à ce jour, les agents ne peuvent pas avoir accès aux prédictions détaillées de l'IA *via* l'interface de codage. Ils ont accès à la cause initiale prédite et au modèle utilisé mais pas aux causes associées prédites.

3.2 Répartition des modes de codage

Au total, les décès survenus en 2023 (codage des causes multiples et détermination de la cause initiale) ont été codés à 62 % par codage automatique à l'aide du système expert Iris/Muse[5], à 15 % par codage manuel assisté par Iris/Muse et à l'aide d'une prédiction par *deep learning* possiblement combinée au système expert Iris/Muse pour les 23 % restant (Tableau 3).

Tableau 3. Effectifs et répartition des certificats reçus en 2023, 2022 et 2021, par mode de codage

	2023		2022		2021	
	Effectif	**Part**	**Effectif**	**Part**	**Effectif**	**Part**
Codage par système expert Iris/Muse	399 000	62 %	415 000	61 %	406 000	61 %
Codage par *IA*	133 000	21 %	16 0000	24 %	149 000	23 %

[5] Version du logiciel Iris : Iris 5.8.1 ; Version du moteur Muse utilisé dans Iris : Muse 2.9 SPECV2021SR40

Codage manuel	95 000	15 %	87 000	13 %	93 000	14 %
dont tirages aléatoires	*47 000*	*7 %*	*44 000*	*7 %*	*52 000*	*8 %*
dont reprise ciblée IA	*41 000*	*6 %*	*35 000*	*5 %*	*21 000*	*3 %*
dont décès sensibles	*3 000*	*0 %*	*3 000*	*0 %*	*3 000*	*0 %*
dont vérifications	*4 000*	*1 %*	*5 000*	*1 %*	*1 0000*	*2 %*
dont codage manuel en routine avant stratégie d'allocation					*7 000*	*1 %*
Trous de collecte	13 000	2 %	14 000	2 %	13 000	2 %
Total	640 000	100 %	676 000	100 %	661 000	100 %

Note de lecture : En 2023, 399 000 certificats ont été codés automatiquement par le système expert Iris sans avoir besoin de vérifications, soit 62% des certificats de cette année.
Champ : ensemble des décès survenus sur le territoire français, en 2021, 2022 et 2023
Sources : CépiDc, bases des décès survenus en 2021, 2022 et 2023

Parmi les certificats de décès 2023, 95 000 certificats ont été codés manuellement, 399 000 codés par le système expert Iris/Muse et 133 000 par IA. Parmi les 95 000 certificats codés manuellement, 47 000 proviennent de tirages uniformes dans la population des certificats de décès de l'année et environ 48 000 de ciblages ou de vérifications (3 000 décès sensibles, 41 000 ciblages liés à l'IA, 4 000 certificats en fin de campagne repris ou codés, ainsi que les vérifications conduisant à des modifications manuelles).

3.3 Allocation des certificats vers la reprise manuelle ou l'inférence par RNP

Environ 41 000 certificats ont été ciblés par une méthode d'IA pour être envoyés en reprise manuelle sur la cause initiale ou les causes associées. La reprise sur la cause initiale a été priorisée de façon à assurer une précision de 94 %, 95 % … puis finalement 97 % pour les causes initiales de chaque catégorie de la *shortlist* européenne [Eurostat, 2012]. Ceci a été réalisé en codant manuellement dans chacune de ces catégories les certificats pour lesquels l'indicateur de confiance dans la prédiction IA de la cause initiale de ce certificat était le plus bas (Tableau 3). Au total, en prenant en compte l'ensemble de la population des décès de l'année quel que soit le mode de codage retenu, on estime qu'après cette reprise ciblée on atteint ou dépasse 97 % de précision en comparant à une campagne traditionnelle (n'utilisant que le système expert Iris/Muse et du codage manuel assisté) pour chacune des catégories de la *shortlist*. Autrement dit, la catégorie de la *shortlist* européenne de laquelle relève le code CIM-10 de la cause initiale ainsi obtenu est la même que celle qui aurait résulté d'une campagne de codage traditionnelle, pour au moins 97 % des certificats codés dans une catégorie donnée de la *shortlist* européenne.

Tableau 4. Nombre de certificats repris manuellement après un ciblage IA sur la cause initiale lors de la campagne de codage 2023, selon le niveau de précision à atteindre pour chaque catégorie de la shortlist européenne

Niveau de précision cible pour chaque catégorie de la *shortlist* européenne	Nombre de certificats ciblés par IA pour une reprise manuelle
94 %	9 519
95 %	4 599
96 %	7 743
96,5 %	5 907
97 %	6 686
Ensemble	34 454

Note de lecture : pour atteindre 94 % de précision dans chacune des catégories de la shortlist *européenne, on*

estime qu'il faut coder manuellement les 9 519 certificats ciblés par l'IA qui sont ceux pour lesquels les catégories en dessous du seuil de 94 % de précision correspondent aux scores de confiance les plus bas.
Champ : ensemble des décès survenus sur le territoire français, en 2023, post-synchronisation Insee
Sources : CépiDc, bases des décès survenus en et 2023

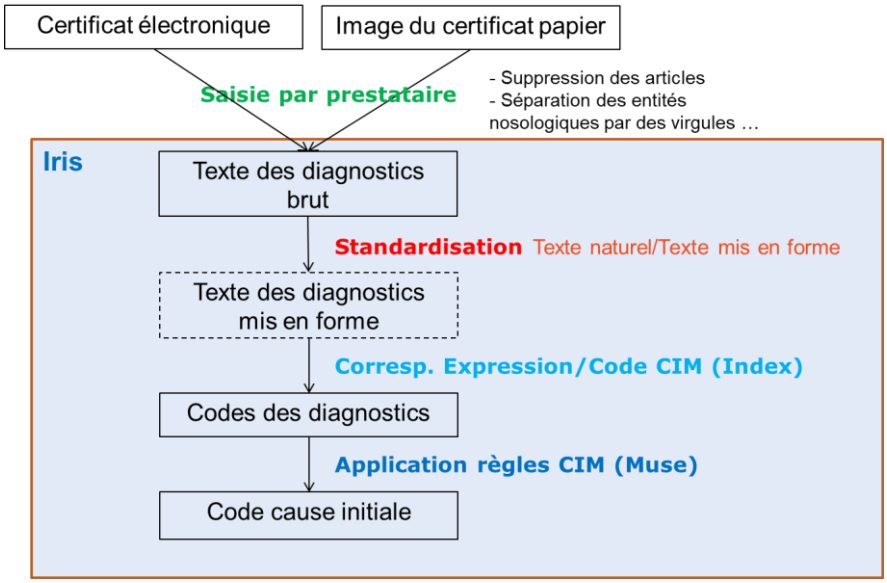
La méthode de ciblage utilisée en 2023 est la même que celle retenue depuis 2021. En première étape, on utilise le modèle mis en œuvre pour le codage final IA de la production du millésime précédent pour obtenir une première prédiction de causes pour l'ensemble des certificats de l'année en cours non codés automatiquement. Puis on calcule pour chacun de ces certificats un indicateur de confiance associée à cette prédiction conditionnellement aux caractéristiques de ce certificat. Cet indicateur mesure la probabilité estimée de cohérence entre la cause initiale prédite par le modèle de *deep learning* et la cause initiale qui aurait été codée dans le cadre d'une campagne de codage traditionnelle (codage par système expert Iris/Muse et codage manuel) : plus il est élevé (proche de 1), plus on considère que la cause initiale prédite a de fortes chances d'être la bonne (Annexe 8.4). Cet indicateur dérive d'un modèle de probabilité linéaire qui explique l'égalité entre les codes CIM-10 de la cause initiale codée par l'équipe de codage et celle prédite par le modèle IA [Zambetta *et al.*, 2023]. Ce modèle n'a pas été réestimé pour la campagne 2023, c'est le même que celui utilisé dans le cadre de la campagne 2022.

En plus des reprises sur les causes initiales, les causes associées des décès de l'année 2023 ont aussi été fiabilisées par un travail de reprise manuelle. Le ciblage se fait cette fois à partir d'une prédiction d'un score BLEU adapté à notre contexte (annexes 8.4.2 et 8.10.1). Pour l'identification des certificats à envoyer en reprise, on se concentre sur les scores les plus faibles pour augmenter le taux de qualité des causes associées indépendamment de la classification Eurostat. En revanche, le nombre de certificats à reprendre sur les causes associées n'a pas été défini comme pour les causes initiales sur une estimation d'une cible de précision mais sur la capacité de traitement de l'équipe de codage. L'objectif de cette reprise est de cibler les certificats au niveau de la qualité des causes associées au lieu de tirer aléatoirement des décès non traités par le système expert. Ainsi, 10 296 certificats divisés en quatre parts égales ont été désignés pour passer par une reprise. Cette reprise arrivant à la fin de la campagne de codage, le dernier lot de 2 576 certificats n'a pas été traité par l'équipe de codage.

3.4 Codage par système expert Iris/Muse

La majorité des certificats sont codés automatiquement en appliquant le système expert de règles international Iris/Muse (Figure 3). Pour les décès 2023, 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].

Figure 3. Codage par système expert Iris/Muse



Les textes des diagnostics issus des certificats électroniques et papier bruts, saisis ou corrigés par le prestataire de saisie et numérisation sont utilisés en entrée du logiciel Iris. Une étape de standardisation faisant intervenir des règles d'expressions régulières est réalisée sur ce texte brut afin de tenter de trouver une correspondance entre le libellé ainsi standardisé (« mis en forme ») et un code CIM-10 à l'aide de l'index (ou « dictionnaire ») implémenté dans Iris et maintenu par le CépiDc. En cas de réussite, les codes de diagnostics sont interprétés par le moteur Muse qui applique les règles internationales d'identification de la cause initiale pour fournir le code qui sera *in fine* retenu en cause initiale. En 2023, le système Iris/Muse conclut à une cause initiale pour 63,7% des certificats de décès. La majorité des certificats non codés sont rejetés à la première étape car le texte ne correspond pas à un libellé du dictionnaire (Pour en savoir plus : « 'Statistiques sur les causes de décès de A à Z »).

3.5 Codage et reprise par l'équipe de codage

L'équipe de codage code en interactif avec le logiciel Iris/Muse et est organisée en trois niveaux.

1) Les codeurs
   Ils prennent en charge le codage de premier niveau, lorsque celui-ci n'est pas automatiquement effectué par le logiciel de codage Iris. Ils maitrisent les règles de codage standard ainsi que les principales expressions médicales, ce qui leur permet de traiter efficacement la majorité des certificats rencontrés.

2) Les nosologistes
   Ce sont des codeurs expérimentés, appelés à intervenir sur les cas les plus complexes. Ils réalisent un codage nécessitant une plus grande expertise et contribuent à l'enrichissement du dictionnaire des termes médicaux issus des certificats. Leur rôle est essentiel pour garantir la qualité et la précision du codage dans les situations complexes.

3) Les experts
   Dotés d'une double compétence en codage et en médecine, les experts interviennent en ultime recours sur les dossiers présentant une complexité clinique majeure. Leur expertise

s'appuie sur une connaissance approfondie de la physiopathologie et des classifications, leur permettant de statuer sur des cas rares ou complexes nécessitant une expertise médicale.

L'équipe de codage traite différents types de certificats : les certificats ciblés à cause d'un faible score de confiance, des certificats tirés aléatoirement dans l'ensemble des certificats non traités par le système expert ainsi que d'autres d'intérêt pour la santé publique et la recherche. Au total, l'équipe de codage a codé ou modifié lors des vérifications environ 95 000 certificats sur les 104 000 passés entre leurs mains (Tableau 5).

Tableau 5. Répartition des codages manuels lors de la campagne 2023, selon l'origine du certificat

Origine du certificat	Effectif
Tirage aléatoire de 100 lots de production complets	24 605
Echantillon démographique permanent	10 116
Aléatoire	11 831
Décès sensibles	2 760
Ciblage IA cause initiale	26 623
Ciblage IA causes associées	7 877
Vérifications IA menant à recodage	3 971
Vérifications batch menant à recodage	333
Ensemble	94 646

Note de lecture : parmi les 94 646 certificats de la base finale codée manuellement, 24 605 sont issus d'un tirage aléatoire dans l'ensemble des certificats non codés par système expert.
Champ : ensemble des décès survenus sur le territoire français, en 2023, post-synchronisation Insee
Sources : CépiDc, bases des décès survenus en 2023

Les décès « sensibles » concernent des décès pour lesquels le CépiDc doit garantir l'exactitude du codage pour des raisons de santé publique. Les décès considérés comme sensibles correspondent à toutes les mentions de SIDA/VIH sur le certificat, aux morts maternelles ainsi qu'à certains décès d'enfants de moins de 15 ans (Annexe 8.6.2). Par rapport à 2022, l'échantillon des décès sensibles a été généré plus tôt dans la production. La part de décès sensibles codés automatiquement par système expert Iris/Muse est aussi plus faible en 2023, en particulier pour les enfants de moins de 28 jours, l'effectif de l'échantillon « décès sensibles non codés par batch Iris » diminue donc naturellement en 2023 (2 890 pour 2023, contre 3 163 en 2022 et 4 639 en 2021), et celui des certificats vérifiés car relevant de décès sensibles augmente (voir après).

Tableau 6. Décès identifiés comme sensibles en 2023, selon la source de sensibilité

Catégorie de décès sensibles	Nombre de certificats
Moins de 28 jours (0-27 jours inclus)	1 198
Au moins 28 jours – Moins de 15 ans	1 048
Morts maternelles	77
Sida/VIH	567
Total	2 890

Note de lecture : En 2023, 2 890 certificats correspondants à des décès sensibles ont été traités et 567 au titre de la surveillance Sida/VIH
Champ : ensemble des décès survenus sur le territoire français, en 2023, post-synchronisation Insee et hors VMC
Sources : CépiDc, bases des décès survenus en 2023

En sus de ces décès dit sensibles, d'autres certificats ont également été systématiquement repris manuellement, à l'issue de la prédiction finale obtenue par *deep learning* :

- tous les certificats (182) pour lesquels le modèle de *deep learning* identifie une cause de type tuberculose, que ce soit en cause initiale ou associée, et que celle-ci n'est finalement pas retenue comme cause initiale ;
- tous les certificats (210) pour lesquels le modèle de *deep learning* identifie une cause liée à la pharmacodépendance ou à la toxicomanie en cause initiale mais que la cause initiale retenue n'appartient pas à cette catégorie

Les volets médicaux complémentaires non codés automatiquement sont aussi systématiquement codés manuellement, soit 104 volets médicaux.

3.6 Codage IA (Deep learning)

Deux évolutions méthodologiques relatives au codage IA ont été mises en œuvre pour la campagne 2023.

Dans un souci de simplification, l'un des deux réseaux de neurones *transformers* mobilisés précédemment pour prédire une séquence de causes a été abandonné, après étude d'impact. En effet, maintenir deux modèles transformers (très proches finalement en termes de structure, de variables en entrée et de base d'entraînement) était trop lourd en temps de calcul et comme de maintenance, compte tenu du gain observé qui était très faible. L'étape de choix de la cause initiale (classification) a été adaptée pour conserver un support de choix assez large : les trois propositions de causes initiales les plus probables (top 1, 2, 3) issues du modèle transformer sont désormais retenues et non plus la seule cause la plus probable (cf. *infra*).

Le modèle de classification final pour la cause initiale (ou surmodèle) qui prédit la cause initiale à retenir parmi les trois propositions les plus probables issues du modèle transformer et celle proposée par le système de règles appliqué à la séquence de codes prédite par le modèle transformer a été amélioré (variables en entrée enrichies et test de performance d'autres architectures de réseau).

3.6.1 Réseau de neurones *transformer* pour coder la séquence de causes

Le modèle *transformer* de *deep learning* mobilisé dans la production 2023 (modèle dit « K5 ») pour coder la séquence de causes suit la même architecture que celui mobilisé en 2022. Seuls des éléments de synthèse concernant la production 2023 sont repris ici (pour une présentation détaillée de ces types de modèle, voir Zambetta *et al.*, 2023 et Zambetta *et al.*, 2024).

Feature engineering/data pipeline.

Les séquences en entrée du modèle sont les concaténations des textes inscrits sur chaque ligne du certificat séparés par le numéro de la ligne. D'autres variables sont également ajoutées à la séquence. Ces variables additionnelles comprennent systématiquement le sexe, le groupe d'âge, l'année de décès, le type de certificat (électronique ou papier), la version du certificat (modèle 1997 ou 2017), ainsi que les circonstances apparentes de décès, nouvelle variable introduite dans les modèles de certificats 2017 qui permet de mieux repérer certaines causes externes (Figure 4).

Ainsi, la phrase d'entrée est :

Paper-back/elec_certificate CertificateVersion sex agegroup yearofdeath sepLine1 text_ written_on_line_1 sepLine2 text_written_on_line2 … … sepLine7 death circumstances sepUC

La phrase de sortie / l'*output* a la même structure que la phrase d'entrée, simplement les codes en CIM-10 remplacent les textes bruts et les circonstances apparentes de décès ne sont pas répétées. Le code de la cause initiale termine la phrase.

Paper-back /elec_certificate certificateVersion sex agegroup yearofdeath sepLine1 ICDcod11 ICDcod12 sepLine2 ICDcod2 … … sepLine7 sepUC ICDcodeUC

Figure 4. Exemple de séquences d'entrée et de sortie du modèle de neurones transformer pour coder la séquence de causes (modèle dit « K5 »)

input sequence	[*start*] certificatpapier versioncertificat1997 femme age55ans annee2017 lignecause1 arrêt cardio respiratoire lignecause2 épanchement pleural lignecause3 métastases pulmonaires lignecause4 cancer sein lignecause7 mort naturelle causeinitiale [*end*]
output sequence	[*start*] certificatpapier versioncertificat1997 femme age55ans annee2017 lignecause1 r092 lignecause2 j90 lignecause3 c780 lignecause4 c509 lignecause7 causeinitiale c509 [*end*]

Pour être utilisées par le modèle, ces séquences sont découpées en éléments de base ou « *token* ». L'algorithme qui réalise ce découpage est "*Tokenizer*" et il n'a pas été spécifiquement ré-entrainé en 2023. En entrée, il découpe les séquences en mots après une étape de normalisation simple (passage en minuscule, suppression des accents et des caractères spéciaux). En sortie il découpe les séquences en codes CIM-10 unitaires et *tokens* spéciaux. En 2023, le dictionnaire en entrée de K5 comprend 206 852 *tokens* et en sortie 6 359 *tokens.*

Architecture du modèle

L'architecture de *transformers* utilisée ici est similaire à celle de la publication initiale [Vaswani 2017, voir 7. Références]. Elle est de type encodeur/décodeur. Les entrées sont représentées par leur plongement dans un espace vectoriel (*embedding*) de taille finie (512) et par la position des mots dans la phrase (*positional encoding*). L'encodeur des modèles *transformers* applique à la séquence d'entrée plusieurs fois les mêmes couches combinant une modélisation du mécanisme d'attention à plusieurs têtes (qui permet de tenir compte des liens entre les mots) et une couche *feed-forward* complètement connectée, tout cela suivi d'une normalisation. Le décodeur répète aussi ces mêmes couches sur la séquence de sortie en intercalant une modélisation du mécanisme d'attention sur *l'output* de l'encodeur. Chaque groupe de couches se termine aussi par une couche *feed-forward* complètement connectée et une étape de normalisation. La sortie du décodeur passe ensuite par une transformation linéaire et une fonction *softmax* permettant de convertir *l'output* du décodeur en probabilités prédites du mot suivant. Pour le modèle K5, 142 millions de paramètres sont à estimer. L'annexe 8.7 illustre l'architecture de ce type de réseau, précise les hyperparamètres choisis et présente les codes du modèle K5 2023.

Bases d'entrainement

Le modèle est entraîné sur des certificats de décès « annotés », c'est-à-dire pour lesquels on dispose de la séquence des causes multiples codées en CIM-10 et de la cause initiale choisie, avec un niveau de qualité jugé suffisant (Annexe 8.5). Les certificats codés par IA ne font pas partie de la base d'entrainement.

Une première base d'entraînement composée de 5 839 339 certificats a été mobilisée fin octobre 2024 pour les travaux d'amélioration de la production. Elle se compose de :

- l'ensemble des données annotées des années 2011 à 2015 (codage automatique et codage manuel),
- l'ensemble des certificats codés automatiquement (système expert) pour 2016 et 2017 ainsi que 300 000 observations tirées aléatoirement parmi celles codées manuellement pour 2016 et 2017,
- l'ensemble des certificats codés automatiquement (système expert) pour les années 2018 et 2019 ainsi que la moitié des certificats codées manuellement pour ces mêmes années au 8 juin 2023 (en respectant cette proportion quel que soit l'échantillon codé)
- 78 % des certificats codés automatiquement (système expert) et 56 % des certificats codés manuellement pour l'année 2020, toujours tirés aléatoirement
- 96 % des certificats codés automatiquement (système expert) et 40 % des certificats codés manuellement en2021 en date du 8 juin 2023 (hors EDP, laissé en test)
- l'ensemble des certificats 2022 codés automatiquement (système expert), 90 % des certificats en reprise IA (lots P1 à P4), 85 % des décès sensibles, 80 % de l'échantillon 1 sur le tirage aléatoire, 80 % des reprises IA P5 et 25 % de l'échantillon 8 aléatoire sur les rejets restants
- 80 % des certificats 2023 de l'échantillon 1 le tirage aléatoire

La deuxième base d'entraînement qui comporte5 862 326 certificats (+ 22 987 observations par rapport à la première) a été utilisée en février 2025 pour produire des résultats provisoires. S'ajoutent à la première base d'entrainement :

- 90% des échantillons de reprises manuelles de l'IA
- 25% des certificats aléatoirement tirés dans les rejets du codage automatique pour être codé en manuel

La troisième base d'entrainement constituée en mars 2025 est utilisée après la reprise manuelle quasiment complète. Elle comprend en plus des certificats de la deuxième base d'entrainement 90 % des certificats codés manuellement lors des reprises de l'IA n°4 et 5 sur la cause initiale, 90 % des décès sensibles et 70 % des reprises de l'IA sur les causes associées soit environ 19 224 certificats de plus que la première base, et au total 5 881 550 certificats. Cette base d'entraînement permet l'inférence finale des causes de décès des certificats non codés ni manuellement ni par le système expert.

La base d'entraînement est toujours séparée en échantillon d'entraînement proprement dit et échantillon de validation comprenant 20 % de l'ensemble des observations tirées aléatoirement une fois, avant l'entraînement.

Base de test

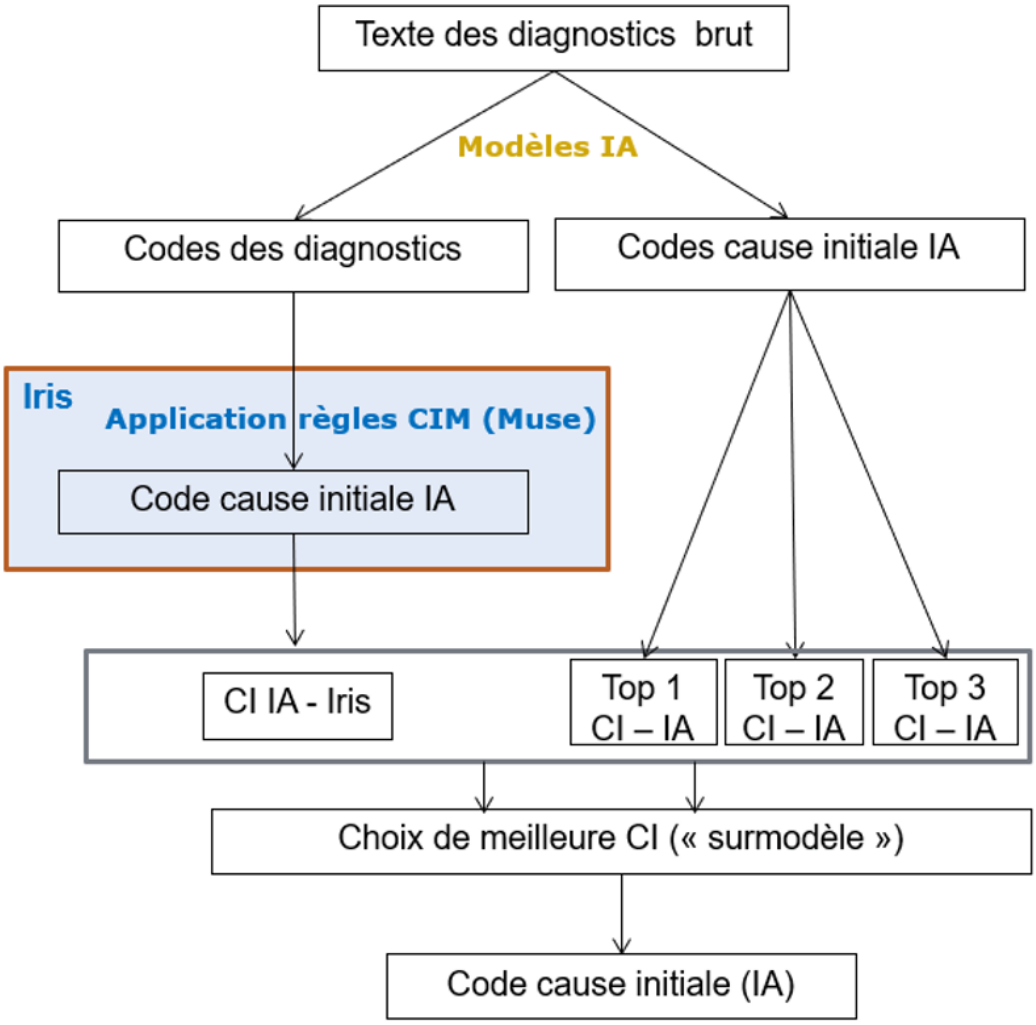
Le test, qui inclut uniquement des observations annotées qui ne figurent pas dans la base d'entraînement, comprend 591 332 observations, dont 418 824 codées manuellement pour des décès de 2016 à 2022.

La table en annexe 8.5 récapitule tous les échantillons et leurs inclusions dans les bases de test et de *train*.

3.6.2 Un sur-modèle pour retenir une cause initiale parmi les différentes propositions possibles

Un modèle de classification en 5 classes est mobilisé pour sélectionner de manière optimale la cause initiale parmi les causes possibles : la prédiction directe de cause initiale obtenue en sortie du modèle K5, proposition la plus probable (K5top 1) mais aussi proposition la deuxième plus probable (K5top 2) et proposition la troisième plus probable (K5top 3), ainsi que la proposition résultant de l'application du système expert Iris/Muse sur la séquence de causes multiples prédites par K5 (irisK5).

Figure 5. Description des propositions possibles de causes initiales impliquant du codage par IA



Ce sur-modèle désigne parmi les types de propositions précédentes celui dont on retiendra la proposition de cause initiale, ou si aucun des modèles n'aboutit à une bonne prédiction. Dans ce dernier cas, on retiendra la prédiction de K5Iris.

L'algorithme retenu pour faire ce choix est un BiLSTM (bidirectionnel long-term short-term memory model) [*Grave et al.*, 2005, Bladi *et al.*, 1999, voir 7. Références]. Il est détaillé dans l'annexe 8.8.

En résumé, les variables explicatives du modèle correspondent aux codes CIM-10 de causes initiales proposées par les différentes classes, les causes multiples, la séquence d'entrée complète du modèle K5 (laquelle concatène les textes médicaux présents sur les certificats à des variables additionnelles d'année, sexe et âge du défunt, type et version du certificat, circonstances apparentes, etc.), ainsi que des variables numériques captant la complexité du certificat (nb de mots, de causes codées…), les probabilités des propositions et leurs écarts (pouvoir discriminant du modèle K5), des indicateurs d'égalité entre le résultat de l'application d'Iris et une des propositions top3 du modèle K5 ou encore sa présence dans les causes multiples (fiabilité de la proposition).

Les variables explicatives n'entrent pas au même moment dans le BiLSTM selon qu'elles sont textuelles ou numériques. Les variables textuelles sont concaténées dans une séquence d'entrée subissant un plongement lexical sur les premières couches du modèle alors que les variables numériques gardent leur caractère numérique.

En termes d'évaluation, le sur-modèle sélectionne le bon « type » de proposition, c'est-à-dire « le modèle dont on devra retenir la proposition de cause initiale » dans 82,9 % des cas dans le test (contre 82,4 % pour la production 2022, 81,6 % pour celle de 2021, 85,1 % pour celle de 2018-2019, la production de ces années antérieures reposent sur deux modèles transformers lesquels n'avaient pas été entraînés sur la même base d'entrainement), et 86,1 % sur les données d'entraînement. Au niveau du code CIM-10 de la cause initiale (c'est-à-dire le code CIM-10 de la cause initiale proposé par la classe/origine retenue), il est correct dans 81,8 % des cas (sur la population test de référence, voir partie évaluation) pour 87,1 % en 2022, 81,5 % en 2021, et 81,9 % en 2018-2019. Ce ne serait le cas que dans 80 % des cas si l'on utilisait seulement le modèle K5Iris, soit un gain de 1,7 point. Le score sur les données d'entraînement est généralement plus élevé que celui du test, car le modèle a vu et appris à partir de ces données. La différence de performance entre le train et le test pourrait indiquer un léger surapprentissage du modèle.

3.7 Vérifications

Dans le processus de codage, la phase de vérification consiste à s'assurer de la qualité du codage de certains certificats lorsque celui-ci apparaît comme « potentiellement incohérent » ou quand ces certificats concernent des décès à fort enjeu en termes de santé publique. Différents types de vérifications ont été effectuées pour l'année 2023, quel que soit le type de codage (manuel, automatique par système expert ou IA). 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. Cette phase de vérification pour l'année 2023 est décrite dans les paragraphes ci-dessous. Un certificat qui a été modifié dans le cadre des vérifications correspond finalement à un certificat codé manuellement, un certificat qui n'est pas modifié gardera son mode de codage initial, ce premier codage étant ainsi confirmé.

En 2023, environ 8 400 certificats ont ainsi été vérifiés : 1 800 initialement codés par l'IA, 3 843 par Iris/muse et 2 718 manuellement (chiffres avant synchronisation avec l'Insee). Suite aux vérifications, 32 % des certificats vérifiés changent de cause initiale : 39 % des certificats codés par IrisMuse ; 22 % des certificats codés manuellement et 32 % des certificats codés par l'IA. Il peut aussi y avoir eu des modifications concernant les causes associées qui n'ont pas conduit à une modification de cause initiale, cela a été le cas pour 3 375 certificats soit 40 % du total. Les évolutions des vérifications par rapport à 2022 sont précisées dans l'annexe 8.6.1.

3.7.1 Vérification de cohérence

Une première série de vérifications consiste à vérifier la cohérence entre le code de cause initiale et le sexe ou l'âge du défunt (cancer de la prostate chez une femme, pathologie hormonale à un âge trop ou pas assez avancé, etc.). Parfois, les incohérences signalées se résolvent sans qu'il soit nécessaire de modifier la cause initiale ou les causes associées. Par exemple ; certaines pathologies, comme un code R54 : Sénilité associé à une personne âgée de moins de 65 ans, génère une anomalie selon les règles. Après vérification du certificat, la CI ou la CA peuvent être confirmées en fonction des informations présentes. D'autres vérifications concernent certains codes qui ne peuvent pas être utilisés en cause initiale de décès selon les recommandations de l'OMS, ils doivent donc être modifiés. Ces vérifications ont concerné 118 certificats : 84 % des causes initiales et 43 % des causes associées ont été modifiées à l'issue de la vérification (pour 161 certificats et 95 % de causes initiales modifiées en 2022).

3.7.2 Décès spécifiques

L'OMS souligne l'importance de garantir un codage optimal pour certains types de décès considérés comme sensibles ou spécifiques, en raison des enjeux majeurs qu'ils représentent en matière de santé publique : les enfants, VIH, morts maternelles (WHO, 2012). La majorité de ces certificats (2 890 avant synchro) figure dans l'échantillon des décès sensibles décrit précédemment et ont fait l'objet d'un codage par un expert. On s'intéresse ici aux autres, ceux codés par système expert ou non encore identifiés lors de la constitution de l'échantillon des décès sensibles. On détaille dans le tableau 7 les certificats concernés ainsi que la part d'entre eux pour lesquels la vérification a conduit à un changement de codage de cause initiale au niveau le plus détaillé de la CIM-10. Dans la stratégie de codage de l'année 2023, les trois vérifications sur les mentions de Covid ont été retravaillées (8.6), en prenant en compte le bilan des années précédentes, pour se concentrer sur les cas méritant un contrôle humain. Ce travail explique la baisse du nombre de certificat à vérifier pour les décès sensibles, pour ces trois vérification (Certains certificats spécifiques avec mention de Covid-19, Les certificats ayant une mention de covid long ou post covid et Les certificats ayant une mention de vaccination de Covid-19) concernaient 3 342 certificats en 2022 et 1 218 pour 2023. A noter que les proportions de causes initiales modifiées pour les cas de décès d'enfants entre 0 et 27 jours (dont les enfants nés sans vie) non codés dans l'échantillon des décès sensibles et les décès d'enfants entre 28 jours et 15 ans avec une mort violente ou pathologie spécifique (dont codes P non codés dans l'échantillon des décès sensibles sont en dessous de 13% indiquant que le codage automatique est plutôt performant sur ces cas.

Il s'agit également, dans cette étape de vérification de repérer les cas de Covid long, de post covid ainsi que ceux en lien avec la vaccination contre la covid-19. On observe une nette diminution, logique dans le contexte de sortie de pandémie, du nombre de certificats inclus dans cette vérification : 543 en 2023 contre 1 447 en 2022. A noter que ces chiffres n'incluent plus les mentions du type « non vacciné contre la covid-19 ».

Tableau 7. Description et nombre de certificats vérifiés pour les décès spécifiques et part de cause initiale modifiées

Description	Nbre de certificats vérifiés	% de CI modifiées	% de CA modifiées
Certains certificats spécifiques avec mention de Covid-19 (cf. 8.6.3) (vérification de l'application de règles complexes de codage, encore récentes)	675	16 %	39 %
Les certificats ayant une cause initiale de la catégorie intention indéterminée (commençant par Y1 ou Y2 ou Y3[0-4])	1 775	11 %	16 %
Les certificats ayant une mention de covid long ou post covid	467	6 %	33 %
Les certificats ayant une mention de vaccination de Covid-19	76	7 %	58 %
Les certificats avec une mention de suicide mais non retenu en cause initiale	107	54 %	57 %
Décès d'enfants entre 28 jours et 15 ans avec une mort violente ou pathologie spécifique (dont codes P non codés dans l'échantillon des décès sensibles (cf. 8.6.2)	493	8 %	11 %
Certificats avec mention de SIDA/VIH non codés dans l'échantillon des décès sensibles	201	9 %	30 %
Décès d'enfants entre 0 et 27 jours (dont les enfants nés sans vie) non codés dans l'échantillon des décès sensibles	579	13 %	23 %
Certificats de morts maternelles non codés dans l'échantillon des décès sensibles	60	8 %	8 %
Les certificats avec une mention d'homicide mais non retenu en cause initiale	36	33 %	44 %
Age entre 28 jours et 6 mois inclus et mention de code P non codés dans l'échantillon des décès sensibles	2	50 %	50 %
Age > 6 mois et mention de code P non codés dans l'échantillons des décès sensibles	5	80 %	60 %
Total	4476	13 %	24 %

Note de lecture : En 2023, 675 certificats vérifiés portaient une mention de COVID-19. Pour 16 % d'entre eux, la cause initiale a été modifiée lors de la vérification et les causes associées changées dans 39 % des cas.
Champ : ensemble des certificats codés avec au moins un code correspondant à un décès spécifique pour vérification manuelle de ces codes
Sources : CépiDc, bases des décès survenus en 2023 avant synchronisation

3.7.3 Difficultés de codage connues liées au logiciel Iris

Le système expert IRIS/MUSE qui permet le codage par l'application d'un système de règles, le codage manuel semi-assisté et une partie du codage réalisé avec des algorithmes de *deep learning* a des lacunes identifiées. Certains certificats ont donc été vérifiés pour pallier ces lacunes. De plus, suite à un choix de privilégier l'amélioration du taux de codage automatique par système expert (bas en France, en comparaison aux autres pays utilisateurs du système expert), des vérifications ont été mises en place sur des certificats dont on sait que le codage peut être erroné. Par exemple : la standardisation du terme « ou » en « , » permet de coder automatiquement et de façon correcte la majorité des certificats, mais une partie, notamment en lien avec les morts violentes doit être vérifiée pour s'assurer que l'interprétation du « ou » est correcte selon les règles de codage. De même avec le terme « sur » qui peut être considéré comme une relation causale (Tableau 7).

Tableau 8. Vérifications liées au système expert, par motif et parts de cause initiale et causes associées modifiées

Réf	Description	Nbre de certificats vérifiés	Part de CI modifiées	% de CA modifiées
M10	Mention de « sur » suivi de certains codes laissant à penser que le « sur » ne doit pas être considéré comme un lien de causalité, avec impact potentiel sur la CI	936	89 %	94 %
M2	Cause initiale en Y avec un code T sur le certificat : dans ces cas-là, le plus souvent une autre pathologie doit être en CI	543	58 %	12 %
M1a	Vérification des certificats avec un code D00- D45 ou D48 en CI alors qu'il y a mention de cancer	265	44 %	52 %
M1b	Vérification des certificats avec un code D46 ou D47 en CI alors qu'il y a mention de cancer	108	55 %	41 %
M11	Anémies en cause initiale (D619, D649 et D539) avec un code informatif (R%%, Z%%, I46%,I269,I10%,I776,I50) présent en partie 1 autre qu'un code d'anémie	286	28 %	39 %
M7	Traitement du "ou" si mention de mort violente à côté du "ou"	194	60 %	44 %
M5	Notion de surdosage AVK en partie 1	84	38 %	45 %
M4	Code d'affection hématologique (D50-D89) en cause initiale alors qu'il existe un code C sur le certificat	152	74 %	59 %
M6	Codes en CI qui commencent par :'M60'ou 'M79' ou 'N94' ou 'N93'' ou 'N92' : Muse s'arrête sur ces codes même s'il y a d'autres codes sur les lignes suivantes. Abandon en 2022 des cas où ces codes barrières étaient sur la dernière ligne de la P1 ou en P2	19	16 %	32 %
M8	D611 en CI et absence de cancer dans le certificat	21	81 %	19 %
M9	Cause initiale à coder en C80.9 si C34.9 ou un C41.1 en cause associée	1	100 %	0 %
Total		2609	65%	56%

Note de lecture : En 2023, 936 certificats ont été vérifiés car il y avait un doute quant à l'interprétation du mot « sur ». La cause initiale a été modifiée dans 89 % des cas et les causes associées dans 94 % des cas.
Champ : ensemble des certificats codés par Iris dont le résultat mérite une vérification manuelle de ces codes à cause des lacunes connues du systèmes
Sources : CépiDc, bases des décès survenus en 2023 avant synchronisation

3.7.4 Vérification de la bonne application des nouvelles règles de 2016

L'OMS a introduit des changements dans les règles de sélection de la cause initiale depuis 2016. Selon la complexité de ces changements, des vérifications sont encore réalisées pour s'assurer de la bonne compréhension et application de ces règles. En 2023, cela concerne, les hépatites virales et certaines chutes.

Tableau 9. Vérifications liées aux nouvelles règles, par motif et parts de cause initiale et causes associées modifiées

Réf	Description	Nbre de certificats vérifiés	Part de CI modifiées	% de CA modifiées
NR1	Les hépatites virales ne peuvent plus être dues à d'autres pathologies : les certificats avec mention d'hépatites virales non retenues en cause initiale	241	16 %	44 %
NR9	Les accidents (V01-X59) ou chutes W00-W19 doivent être dû à une autre cause	6	50 %	33 %
Total		247	17 %	43 %

Note de lecture : En 2023, 241 certificats ont été vérifiés car liés à une hépatite virale. La cause initiale a été modifiée dans 16 % des cas et les causes associées dans 44 % des cas.
Champ : ensemble des certificats codés avec au moins un code correspondant à l'une des nouvelles règles de l'OMS pour vérification manuelle de ces codes
Sources : CépiDc, bases des décès survenus en 2023 avant synchronisation

3.7.5 Vérifications liées aux évolutions du dictionnaire

A partir de 2018, le dictionnaire en entrée du système expert a été mis à jour, notamment les libellés faisant l'objet de « choix de code ». On définit un « choix de code » comme étant un libellé qui peut être codé de plusieurs façons différentes selon le contexte du certificat. En l'état, un regard humain est systématiquement nécessaire, alors que, dans certains cas, le choix opéré *in fine* est toujours le même ou qu'un code n'est plus pertinent. Des suppressions de choix de codes au profit d'un code « préféré » ont été réalisées et ont permis de diminuer le nombre de certificats inutilement codés manuellement. Cependant, certaines situations peuvent laisser le doute et des vérifications ont donc été définies. Par exemple, la présence du libellé hémorragie intracérébrale pouvait être codé soit en S068 (hémorragie intracérébrale traumatique) soit en I6199 (hémorragie intracérébrale non précisée) : le code S068 a été supprimé, pour permettre le codage automatique en I6199. Cependant, il a été décidé de vérifier les certificats qui mentionne une hémorragie intracérébrale lorsque celle-ci est en présence d'une cause externe. En effet, sur ces certificats, il est possible que l'hémorragie intracérébrale soit alors codée en S068. Le contexte du certificat permet au codeur, lors de sa vérification de statuer et d'éventuellement recoder l'hémorragie intracérébrale correctement.

Les vérifications liées aux choix de code ont été recalibrées à partir de 2022, ce qui explique leur ampleur nettement réduite par rapport à 2021 (678 en 2023, 727 en 2022 contre 2 958 en 2021). Certaines vérifications, jugées peu pertinentes car entrainant trop peu de modifications de causes initiales, ont été supprimées ou mieux ciblées (Tableau 10).

Tableau 10. Vérifications liées à la démarche de choix de code, par motif et parts de cause initiale et causes associées modifiées

Réf	Description	Nbre de certificats vérifiés	Part de CI modifiées	% de CA modifiées
CC_Hemorragie_intracerebrale	S'il y a une notion de causes externes dans un certificat avec mention d'hémorragie intracérébrale, le codage sera S068 et non I6199	99	24 %	65 %
CC_Alcool	S'il y a une notion de cause externe dans un certificat avec mention d'alcool, on vérifie que le code F10.2 est approprié	195	14 %	57 %
CC_choc_hemorragique_hypovolemique	S'il y a une notion d'acte diagnostique et thérapeutique invasif dans un certificat avec mention de choc hémorragique ou hypovolémique, on vérifie que le code R57.1 est approprié (potentiellement à reprendre en T179). En 2022 n'est plus réalisée sur le codage manuel	151	26 %	93 %
CC_IRA	On vérifie que le code J96.0 est approprié pour les certificats avec un libellé « IRA » et une notion de diabète ou certaines pathologies rénales/urinaires.	179	23 %	85 %
CC_Rhabdomyolyse	S'il y a une notion de traumatisme ou cause externe (2022 : codes T ou S et Cause initiale différente de X) dans un certificat avec mention de rhabdomyolyse, on vérifie que le code M62.8 est bien approprié.	23	17 %	65 %
CC_noyage	Libellé « noyade » et une information de dépression (F30, F32 ou F33) en partie 1 si les circonstances apparentes du décès sont inconnues.	12	83 %	33 %
CC_paraplegie	On vérifie que le code G822 est approprié pour coder le libellé « paraplégie » s'il y a aussi une mention d'AVC sur le certificat. En 2022, n'est plus réalisée sur le codage réalisé par les méthodes d'IA	19	32 %	89 %
total		678	23 %	75 %

Note de lecture : En 2023, 99 certificats ont été vérifiés car liés à une hémorragie intracérébrale. La cause initiale

a été modifiée dans 24 % des cas et les causes associées dans 65 % des cas.
Champ : ensemble des certificats codés avec au moins un diagnostic donnant lieu à un choix entre plusieurs codes pour vérification manuelle de ces certificats
Sources : CépiDc, bases des décès survenus en 2023 avant synchronisation

3.7.6 Vérifications pour raisons multiples

Certains certificats sont vérifiés pour plusieurs raisons citées ci-dessus. Ces cas de vérifications multiples ont concerné 233 certificats en 2023 et pour 58 % d'entre eux, la cause initiale a été modifiée à l'issue de la vérification.

4 Erreur ! Source du renvoi introuvable.Evaluation de la stratégie de codage

Le passage à une campagne en trois modes de codage n'est pas simplement l'inclusion et le remplacement d'une partie du codage manuel par des prédictions par IA, mais l'instauration d'une véritable stratégie de codage selon les caractéristiques du certificat. Le but de cette stratégie est d'assurer la cohérence du codage dans le temps sur des champs spécifiques (décès sensibles, champ EDP) et l'utilisation la plus efficace possible de l'IA et de l'expertise humaine *via* le codage manuel.

L'objectif de cette partie est d'étudier la performance (cohérence) de cette stratégie au regard d'un contrefactuel artificiel créé à partir d'une population test de référence représentative de la distribution des causes de décès dans la population. Ce qui est évalué ici n'est pas la qualité de l'imputation par IA en tant que telle, mais le niveau de performance de la stratégie atteint, au sens de la cohérence entre le codage qui aurait été obtenu dans une campagne classique de codage combinant codage par système expert et codage manuel assisté et celui obtenu en 2023. *In fine*, l'évaluation permet notamment d'identifier les variations de cohérences liées aux décisions structurelles de la campagne (nouveau surmodèle, abandon d'un modèle) et celles liées aux décisions d'orientation des certificats (codage manuel ou prédiction par IA) au cours de la campagne.

4.1 Population test de référence

La population test de référence[6] est composée de 958 679 observations, représentatives de la répartition des causes de décès pour chacune des années 2016, 2017, 2020, 2021, 2022 et 2023. La proportion de codage automatique par système expert dans cette population est de 59,2 %, ce qui est un peu moins que le taux de codage automatique effectif par ce biais en 2023 (63,5 %, hors trous de collecte). La conséquence sera donc une légère sous-estimation de la cohérence du codage dans les tableaux présentés ci-dessous.

On simule les apports de la reprise manuelle aléatoire et ciblée telle qu'elle a été menée en 2023. Le raisonnement est le suivant : parmi les 233 317 certificats non codés automatiquement par système expert en 2023[7], 46 872 certificats répartis en 4 échantillons ont été sélectionnés aléatoirement et codés manuellement, soit 20,1 % de l'ensemble des certificats non codés automatiquement par système expert. On va donc simuler, *via* un tirage uniforme, que 20,1 % des certificats à coder manuellement dans le test vont passer en reprise manuelle. Pour les décès sensibles, on applique les règles de repérage utilisées pendant la campagne pour les identifier. Pour simuler l'impact de la reprise ciblée IA, on s'appuie sur les scores de confiance issues du modèle « K5 » et les prédictions de cause initiale de décès de « K5Iris ». On applique la même part de reprise dans les catégories sur lesquelles la reprise manuelle a été ciblée que ce qui a été fait sur 2023. Nouveauté de la campagne 2023, on simule également la reprise ciblée IA spécifique à l'entièreté du processus morbide. On s'appuie sur les scores BLEU de la transcription par l'IA, et on prélève la même proportion que pendant la campagne. Les vérifications, qui constituent la dernière étape de la campagne, sont simulées. Cependant, ne sont pas simulés les 4 162 certificats codés manuellement alors qu'ils ne relevaient pas

[6] Voir l'annexe 8.11 pour la constitution de la population test de référence.

[7] Les répartitions utilisées sont basées sur les estimations pendant la campagne, non pas celles des données finales. C'est un point d'amélioration pour la prochaine itération. Il ne tient pas compte de la synchronisation notamment.

d'un échantillon spécifique (vérifications manuelles de certains certificats et plus rarement du fait d'un blocage à l'intégration).

La répartition des types de codage dans la campagne 2023 et dans sa simulation sur la population test de référence montre que le taux de codage automatique est légèrement inférieur dans la simulation, comme attendu (Tableau 11).

Tableau 11. Description de la distribution des allocations de codage dans la base 2023 et sa simulation dans la population test de référence

	Base 2023		**Population test de référence**	
	Effectif	**Répartition (%)**	**Effectif**	**Répartition (%)**
Codage automatique	398 144	62,3 %	567 225	59,2 %
Reprise manuelle aléatoires (y compris EDP)	45 800	7,2 %	45 905	8,2 %
Manuel EDP	9 969	1,6 %	78 815	1,8 %
Décès sensible	4 108	0,6 %	6 164	0,6 %
Reprise ciblée IA CI	33 153	5,2 %	47 705	5,0 %
Reprise ciblé IA CA	7 877	1,2 %	13 831	1,4 %
Vérifications	134	0,0 %	1 358	0,1 %
Total Reprise manuelle	95 212	14.9 %	171 195	17,9 %
Total prédiction IA	133 341	20,8 %	246 509	25,7 %
Certificats imputés	12 850	2,0 %		
Total*	**639 547**	**100,0 %**	**958 679**	**100,0 %**

Lecture : Le taux de décès sensible dans la base de test est de 0,6%, soit la même proportion que pour l'année 2023. Le taux de codage automatique est de 59,2% dans la base de test, contre 62,3% pour l'année de 2023.
Champ : Ensemble des décès synchronisés de 2023 et Population test de référence composée des décès déjà codé des années 2016, 2017, 2020, 2021, 2022 et 2023
Source : CépiDc, bases des décès survenus en 2023 et base de test

4.2 Accuracy, cohérence globale et comparaison à la campagne précédente

Sur la partie de la population test de référence qui aurait été codée manuellement dans une campagne classique de codage, la cause initiale obtenue en combinant la prédiction du sur-modèle et la reprise manuelle ciblée est égale à la cause initiale codée par l'équipe de codage au niveau le plus fin de la CIM-10 dans 91,6 % des cas (contre 90,0 % sur la campagne 2022). Elle relève de la même catégorie de la *shortlist* européenne dans 94,8 % des cas (pour 93,7 % en 2022). Les tableaux 12 et 13 détaillent les performances des prédictions par les modèles d'IA, combinés ou non avec Iris/Muse et avec la reprise manuelle ciblée telle qu'elle a été réalisée sur 2023.

Le tableau 12 se concentre sur les certificats qui auraient été codés manuellement dans une campagne traditionnelle ne combinant que batch automatique et codage manuel. Le Tableau 13 reporte les mêmes indicateurs calculés sur l'ensemble de la population test de référence, c'est-à-dire y compris les certificats codés par batch pour lesquels il n'y a pas eu de changement entre cette campagne et

une campagne traditionnelle (cohérence complète), de façon à fournir un niveau global de cohérence entre campagnes prenant en compte tous les modes de codage.

Tableau 12. Cohérence (accuracy) entre causes initiales prédites par deep learning (K5), combinaison de deep learning et Iris Muse (K5 iris), proposition du surmodèle, avec ou sans prise en compte de la reprise manuelle 2023 pour les données qui auraient été codées manuellement dans une campagne classique de codage

Année	K5	K5Iris	Surmodèle	Surmodèle + reprise manuelle	Nombre d'observations dans la base simulée	Surmodèle + reprise manuelle (2022)
***Accuracy* au niveau du code CIM-10**						
Ensemble	**0,792**	**0,803**	**0,817**	**0,916**	**391 454**	**0,900**
2016	0,779	0,785	0,798	0,910	93 144	0,892
2017	0,775	0,784	0,798	0,907	93 912	0,891
2020	0,806	0,819	0,835	0,920	121 461	0,911
2021	0,803	0,819	0,835	0,919	23 520	0,909
2022	0,806	0,826	0,837	0,927	25 491	0,904
2023	0,804	0,822	0,835	0,926	33 926	-
***Accuracy* au niveau de la catégorie de la shortlist européenne**						
Ensemble	**0,860**	**0,867**	**0,878**	**0,948**	**391 454**	**0,937**
2016	0,852	0,854	0,866	0,945	93 144	0,931
2017	0,848	0,852	0,863	0,942	93 912	0,93
2020	0,871	0,878	0,891	0,951	121 461	0,945
2021	0,867	0,877	0,888	0,951	23 520	0,941
2022	0,867	0,881	0,889	0,954	25 491	0,941
2023	0,868	0,880	0,889	0,955	33 926	-

Source : Inserm – CépiDc.

Lecture : Au niveau le plus fin de la Classification Internationale des Maladies, la première prédiction du modèle K5 est correcte dans 79,2% des cas qui auraient été codés manuellement dans une campagne traditionnelle. Si IRIS/Muse propose une cause initiale à partir des causes associées proposées par K5, on passe à 80,3% de cohérence. Si on utilise le surmodèle pour départager les trois prédictions proposées par K5 et celle proposée par Iris/Muse, on obtient 81,7% de certificats codés avec le même code CIM-10. En prenant en compte toutes les étapes jusqu'à la reprise manuelle, on arrive à une accuracy de 91,6%.

Champ : Population test de référence - uniquement les données codées manuellement dans le cadre d'une campagne classique.

Source : CépiDc, bases des décès survenus en 2023 et base de test

Au niveau le plus fin de la CIM-10, la prédiction du modèle K5 est correcte dans 79,2 % des cas (79,0 % en 2022). Appliquer Iris/Muse sur la séquence des causes prédites par K5 lorsqu'il obtient une réponse non ambiguë fait gagner un point de cohérence. Le surmodèle qui choisit entre les trois meilleures propositions de cause initiale par K5 et K5Iris (K5 combiné avec Iris/Muse) permet de gagner 1,4 point supplémentaire. La prise en compte de la reprise aléatoire et ciblée permet de gagner 9,9 points supplémentaires (contre 8,3 points en 2022) et d'atteindre les 91,6 %. L'évaluation de chaque étape de la reprise ciblée sera détaillée par la suite. Au niveau *shortlist* européenne, le sur-modèle permet de gagner 1,1 point *d'accuracy* par rapport à K5Iris (pour 1,3 point en 2022) et la reprise aléatoire et ciblée 7,0 points (pour 6,0 points en 2022). Au total, on atteint 94,8 % de cohérence (93,7 % en 2022).

La cohérence est relativement homogène sur les cinq années de validité des données. Ce sont les données les plus anciennes de 2016 et 2017 qui ont le taux de cohérence le plus faible, les données récentes sont donc mieux traitées par la campagne actuelle.

Tableau 13. Cohérence entre causes initiales prédites par deep learning (K5), combinaison de deep learning et Iris Muse (K5Iris), proposition du surmodèle, avec ou sans prise en compte de la reprise manuelle 2023 pour l'ensemble de la population (et 2022 pour référence)

Année	K5 + automatique	K5Iris + automatique	Surmodèle + automatique	Surmodèle + reprise manuelle + automatique	Nombre d'observations	Surmodèle + reprise manuelle + automatique (2022)
***Accuracy* au niveau du code CIM-10**						
Ensemble	**0,915**	**0,920**	**0,925**	**0,966**	**958 679**	**0,959**
2016	0,907	0,910	0,915	0,962	221 807	0,955
2017	0,907	0,910	0,916	0,961	226 856	0,955
2020	0,918	0,923	0,93	0,966	285 784	0,962
2021	0,927	0,933	0,939	0,97	63 058	0,966
2022	0,927	0,935	0,939	0,973	67 979	0,964
2023	0,929	0,935	0,940	0,973	93 195	-
***Accuracy* au niveau de la catégorie de la *shortlist* européenne**						
Ensemble	**0,943**	**0,946**	**0,950**	**0,979**	**958 679**	**0,974**
2016	0,938	0,939	0,944	0,977	221 807	0,971
2017	0,937	0,939	0,943	0,976	226 856	0,971
2020	0,945	0,948	0,954	0,979	285 784	0,977
2021	0,950	0,954	0,958	0,982	63 058	0,978
2022	0,950	0,956	0,958	0,983	67 979	0,978
2023	0,952	0,956	0,960	0,984	93 195	-

Lecture : Au niveau de la catégorie agrégée de la shortlist d'Eurostat, la prédiction donnée par le codage automatique ou, à défaut, le modèle K5 est correcte dans 94,3 % des cas qui auraient été codés automatiquement ou manuellement dans une campagne traditionnelle. Si plutôt que K5, IRIS/Muse propose une cause initiale à partir des causes associées proposées par K5, on passe à 94,6 % de cohérence. Si on utilise le surmodèle pour départager les trois prédictions proposées par K5 et celle proposée par Iris/Muse, on obtient 95% de certificats codés avec le même code CIM-10. Si finalement, on récupère la prédiction du codage automatique, ou à défaut celle du codage manuel, ou à défaut celle du surmodèle, on arrive à une accuracy de 97,9 %.
Champ : Population test de référence.
Source : CépiDc, base de test

4.3 *Précision, rappel et écarts d'effectifs*

Les tableaux 14 et 15 présentent les précisions, rappels, F-mesures et les effectifs prédits par catégorie de la *shortlist* européenne, pour le sur-modèle combiné au batch et lorsque l'on tient aussi compte de la reprise manuelle. La précision est la part de prédictions correctes rapportée à l'ensemble des prédictions dans la catégorie, le rappel est la part des observations correctement prédites par le modèle rapporté à l'ensemble des observations réellement dans la catégorie ; la F-mesure est la moyenne harmonique des deux.

Tableau 14. Performances du sur-modèle et du sur-modèle combiné avec la reprise manuelle évalués sur les observations de la population test qui auraient été codées manuellement.

		Surmodèle						Surmodèle + reprise manuelle					
Shortlist EUROSTAT	Effectifs réels (2023)	Effectifs prédits	Précision	Rappel	F-measure	pred/réels - 1	Signif.	Effectifs prédits	Précision	Rappel	F-measure	pred/réels - 1	Signif.
01.1 Tuberculose	477	409	0,907	0,778	0,837	-0,143	****	469	0,979	0,962	0,97	-0,017	
01.2 SIDA (maladie VIH)	309	302	0,768	0,751	0,759	-0,023		308	1	0,997	0,998	-0,003	
01.3 Hépatites virales	369	344	0,721	0,672	0,696	-0,068		339	0,929	0,854	0,89	-0,081	*
01.4 Autres maladies infectieuses et parasitaires	6853	6624	0,8	0,773	0,786	-0,033	****	6645	0,947	0,918	0,933	-0,03	***
02.1.01 Tumeur lèvre, bouche, pharynx	3167	3037	0,926	0,888	0,907	-0,041	***	3138	0,964	0,955	0,96	-0,009	
02.1.02 Tumeur de l'œsophage	2767	2806	0,942	0,955	0,948	0,014		2817	0,963	0,981	0,972	0,018	
02.1.03 Tumeur de l'estomac	2752	2659	0,956	0,924	0,94	-0,034	**	2726	0,972	0,963	0,968	-0,009	
02.1.04 Tumeur Colon, rectum, anus	11364	11240	0,954	0,944	0,949	-0,011		11397	0,972	0,975	0,974	0,003	
02.1.05 Tumeur du foie et voies biliaires	5616	5441	0,952	0,922	0,937	-0,031	***	5539	0,973	0,959	0,966	-0,014	
02.1.06 Tumeur du Pancréas	6281	6214	0,969	0,959	0,964	-0,011		6280	0,981	0,981	0,981	0	
02.1.07 Tumeur du Larynx	725	724	0,89	0,888	0,889	-0,001		733	0,936	0,946	0,941	0,011	
02.1.08 Tumeur du Poumon	18247	18123	0,956	0,95	0,953	-0,007		18291	0,973	0,975	0,974	0,002	
02.1.09 Mélanome malin de la peau	1362	1383	0,915	0,93	0,922	0,015		1389	0,946	0,965	0,955	0,02	
02.1.10 Tumeur du sein	7983	7951	0,955	0,951	0,953	-0,004		7992	0,972	0,973	0,973	0,001	
02.1.11 Tumeur du col de l'utérus	599	593	0,931	0,922	0,926	-0,01		603	0,962	0,968	0,965	0,007	
02.1.12 Tumeur autre partie de l'utérus	1921	1846	0,945	0,908	0,926	-0,039	**	1882	0,968	0,948	0,958	-0,02	
02.1.13 Tumeur de l'ovaire	2035	2009	0,951	0,939	0,945	-0,013		2031	0,971	0,969	0,97	-0,002	
02.1.14 Tumeur de la prostate	5589	5525	0,947	0,936	0,942	-0,011		5583	0,968	0,967	0,968	-0,001	
02.1.15 Tumeur du rein	2463	2396	0,937	0,911	0,924	-0,027	*	2446	0,959	0,952	0,955	-0,007	
02.1.16 Tumeur de la vessie	3416	3395	0,94	0,934	0,937	-0,006		3410	0,966	0,965	0,965	-0,002	
02.1.17 Tumeur cerveau et système nerveux central	2597	2443	0,949	0,893	0,92	-0,059	****	2528	0,98	0,954	0,967	-0,027	*
02.1.18 Tumeur de la thyroïde	296	287	0,899	0,872	0,885	-0,03		286	0,976	0,943	0,959	-0,034	
02.1.19 Maladie de Hodgkin et lymphomes	3752	3747	0,934	0,933	0,933	-0,001		3775	0,965	0,971	0,968	0,006	
02.1.20 Leucémie	4160	4194	0,931	0,939	0,935	0,008		4201	0,961	0,971	0,966	0,01	
02.1.21 Autres Tumeur des tissus lymphoïdes et hématopoiétiques	2258	2227	0,921	0,909	0,915	-0,014		2256	0,957	0,957	0,957	-0,001	
02.1.22 Autres Tumeurs malignes	17522	18631	0,838	0,891	0,863	0,063	****	17672	0,935	0,943	0,939	0,009	
02.2 Tumeurs non malignes	6080	6279	0,825	0,852	0,839	0,033	***	6158	0,932	0,944	0,938	0,013	
03 Maladies du sang et hématopoiétiques	2444	2134	0,719	0,628	0,67	-0,127	****	2342	0,854	0,818	0,835	-0,042	***
04.1 Diabète sucre	8491	8087	0,91	0,867	0,888	-0,048	****	8320	0,956	0,936	0,946	-0,02	**
04.2 Autres maladies endocriniennes, nutritionnelles et métaboliques	7140	6903	0,82	0,792	0,806	-0,033	****	6976	0,931	0,91	0,92	-0,023	**
05.1 Démences	9542	10239	0,856	0,918	0,886	0,073	****	10127	0,901	0,956	0,928	0,061	****
05.2 Abus d'alcool	1835	1911	0,803	0,837	0,82	0,041	**	1821	0,949	0,942	0,945	-0,008	
05.3 Pharmacodépendance, toxicomanie	255	228	0,732	0,655	0,692	-0,106	**	255	0,976	0,976	0,976	0	
05.4 Autres troubles mentaux et du comportement	3027	2969	0,803	0,788	0,795	-0,019		2963	0,931	0,912	0,922	-0,021	
06.1 Maladie de Parkinson	3407	3410	0,92	0,921	0,921	0,001		3419	0,95	0,954	0,952	0,004	
06.2 Maladie d'Alzheimer	9140	9130	0,944	0,943	0,944	-0,001		9192	0,967	0,972	0,97	0,006	
06.3 Autres maladies du système nerveux et des organes des sens	8685	8533	0,836	0,821	0,828	-0,018	*	8634	0,94	0,935	0,937	-0,006	
07.1.1 Infarctus aigu du myocarde	7345	7583	0,877	0,905	0,891	0,032	****	7579	0,924	0,954	0,939	0,032	****
07.1.2 Autres Cardiopathies ischémiques	12772	12429	0,89	0,866	0,878	-0,027	****	12643	0,94	0,931	0,936	-0,01	
07.2 Autres maladies du cœur	27443	27696	0,863	0,871	0,867	0,009	*	27750	0,932	0,942	0,937	0,011	**
07.3 Maladies cérébro-vasculaires	21714	21663	0,893	0,891	0,892	-0,002		21843	0,946	0,951	0,949	0,006	
07.4 Autres maladies de l'appareil circulatoire	17342	16820	0,858	0,833	0,845	-0,03	****	16966	0,947	0,926	0,937	-0,022	****
08.1 Grippe	948	933	0,924	0,909	0,917	-0,016		942	0,969	0,963	0,966	-0,006	
08.2 Pneumonie	5384	5420	0,838	0,844	0,841	0,007		5443	0,915	0,925	0,92	0,011	
08.3.1 Asthme	497	458	0,871	0,803	0,836	-0,078	**	491	0,935	0,924	0,929	-0,012	
08.3.2 Autres maladies chroniques des voies respiratoires inferieures	6705	6845	0,885	0,904	0,894	0,021	**	6821	0,944	0,96	0,952	0,017	*
08.4 Autres maladies de l'appareil respiratoire	7823	7917	0,777	0,787	0,782	0,012		7821	0,905	0,905	0,905	0	
09.1 Ulcère gastro duodénal	712	736	0,842	0,871	0,856	0,034		725	0,931	0,948	0,939	0,018	
09.2 Cirrhoses, fibroses et hépatites chroniques	4748	4766	0,902	0,906	0,904	0,004		4795	0,95	0,96	0,955	0,01	
09.3 Autres maladies de l'appareil digestif	13328	13339	0,852	0,852	0,852	0,001		13350	0,942	0,944	0,943	0,002	
10 Maladies de la peau et du tissu sous cutané	1461	1574	0,74	0,797	0,768	0,077	****	1448	0,936	0,928	0,932	-0,009	
11.1 Arthrite rhumatoïde et ostéoarthrite	518	471	0,837	0,761	0,797	-0,091	***	490	0,959	0,907	0,933	-0,054	
11.2 Autres maladies du système ostéoarticulaire et des muscles	3770	3734	0,726	0,719	0,722	-0,01		3630	0,94	0,905	0,922	-0,037	***
12.1 Maladies du rein et de l'urètre	5191	4844	0,832	0,777	0,803	-0,067	****	4968	0,936	0,896	0,916	-0,043	****
12.2 Autres maladies génito-urinaires	2855	2837	0,812	0,807	0,81	-0,006		2859	0,926	0,927	0,926	0,001	
13 Complications de grossesse et d'accouchement	63	16	0,938	0,238	0,38	-0,746	****	63	1	1	1	0	
14 Certains affections dont l'origine est périnatale	2036	2084	0,93	0,952	0,941	0,024		2039	0,998	0,999	0,998	0,001	
15 Malformations congénitales et anomalies chromosomiques	1608	1266	0,87	0,685	0,766	-0,213	****	1514	0,987	0,93	0,958	-0,058	***
16.1 Syndrome de mort subite du nourrisson	206	210	0,929	0,947	0,938	0,019		207	0,995	1	0,998	0,005	
16.2 Causes inconnues ou non précisées	6060	5991	0,808	0,799	0,804	-0,011		5872	0,945	0,916	0,93	-0,031	***
16.3 Autres symptômes et états morbides non définis	8035	8592	0,815	0,872	0,843	0,069	****	8367	0,909	0,946	0,927	0,041	****
17.1.1 Accidents de transport	2680	2557	0,94	0,897	0,918	-0,046	***	2675	0,979	0,978	0,979	-0,002	
17.1.2 Chutes accidentelles	11080	11373	0,916	0,941	0,928	0,026	****	11117	0,975	0,978	0,976	0,003	
17.1.3 Noyade et submersion accidentelle	475	502	0,833	0,88	0,856	0,057		482	0,952	0,966	0,959	0,015	
17.1.4 Intoxications accidentelles	1939	2156	0,703	0,782	0,74	0,112	****	1923	0,945	0,938	0,941	-0,008	
17.1.5 Autres accidents	16385	16362	0,858	0,857	0,857	-0,001		16071	0,961	0,943	0,952	-0,019	***
17.2 Suicide	6065	6026	0,935	0,929	0,932	-0,006		6118	0,974	0,982	0,978	0,009	
17.3 Homicides	493	459	0,863	0,803	0,832	-0,069	*	484	0,99	0,972	0,981	-0,018	
17.4 Evènement dont l'intention n'est pas déterminée	1850	1865	0,597	0,602	0,599	0,008		1852	0,877	0,878	0,878	0,001	
17.5 Autre cause externe	1820	1928	0,351	0,372	0,361	0,059	***	1684	0,839	0,776	0,807	-0,075	****
18 COVID-19	15177	15629	0,946	0,974	0,96	0,03	****	15479	0,967	0,986	0,976	0,02	***

Note : les degrés de significativité des écarts entre effectifs prédits et réels proviennent de tests d'égalité supposant que les effectifs réels suivent des lois de Poisson - * pval < 0.2 , ** pval < 0.1, *** pval < 0.05, **** pval < 0.01 : voir https://www.cepidc.inserm.fr/documentation/tests-statistiques-relatifs-aux-indicateurs-de-mortalite-en-population

Source : Inserm - CépiDc

Données : Base de données de test et base 2023 - uniquement les données codées manuellement.

Lecture : Dans le cadre de l'évaluation du transfert entre codage manuel et prédiction IA, sur un total de 15177 certificats avec une cause initiale « COVID-19 » à identifier, le surmodèle en identifie 15 629, soit une précision de 94 ,6% (sur 100 certificats que le modèle prédit « COVID-19 », 94,6 sont effectivement des certificats avec cause initiale de décès « COVID-19 ») et un rappel de 97,4% (sur 100 certificats appartenant réellement à la cause initiale de décès « COVID-19 », le surmodèle en prédit correctement 97,4). La F-measure qui combine ces deux proportions, est 0,96, c'est-à-dire la plus élevée de toutes les causes. Le surmodèle est donc très bon pour identifier les certificats par décès « COVID-19 » en cause initiale. Il y a 3% de prédictions de cause initiale « COVID-19 » de trop par rapport aux effectifs réels (c'est-à-dire codés manuellement). Cet écart d'effectifs est significatif, c'est-à-dire que l'écart entre effectifs prédits et effectifs réels a très peu de chance d'être dû au hasard, le surmodèle surévalue en moyenne la quantité de certificats « COVID-19 ». La reprise manuelle, qui correspond à une stratégie de ciblage des certificats les plus complexes pour les rediriger vers le codage manuel, permet d'augmenter la F-measure de 16 points de pourcentage.

Tableau 15. Performances du système expert combiné au sur-modèle et au sur-modèle combiné et à la reprise manuelle évaluées sur l'ensemble de la population test de référence, par item de la shortlist Eurostat.

		Codage automatique + Surmodèle						Codage automatique + Surmodèle + reprise manuelle					
Shortlist EUROSTAT	Effectifs réels (2023)	Effectifs prédits	Précision	Rappel	F-measure	pred/réels - 1	Signif.	Effectifs prédits	Précision	Rappel	F-measure	pred/réels - 1	Signif.
01.1 Tuberculose	531	463	0,918	0,8	0,855	-0,128	****	523	0,981	0,966	0,973	-0,015	
01.2 SIDA (maladie VIH)	392	385	0,818	0,804	0,811	-0,018		391	1	0,997	0,999	-0,003	
01.3 Hépatites virales	626	601	0,84	0,807	0,823	-0,04		596	0,96	0,914	0,936	-0,048	
01.4 Autres maladies infectieuses et parasitaires	15896	15667	0,915	0,902	0,909	-0,014	**	15688	0,978	0,965	0,971	-0,013	**
02.1.01 Tumeur lèvre, bouche, pharynx	5860	5730	0,961	0,939	0,95	-0,022	**	5831	0,981	0,976	0,978	-0,005	
02.1.02 Tumeur de l'œsophage	5695	5734	0,971	0,978	0,975	0,007		5745	0,982	0,991	0,986	0,009	
02.1.03 Tumeur de l'estomac	6908	6815	0,983	0,97	0,976	-0,013		6882	0,989	0,985	0,987	-0,004	
02.1.04 Tumeur Colon, rectum, anus	27459	27335	0,981	0,977	0,979	-0,005		27492	0,988	0,99	0,989	0,001	
02.1.05 Tumeur du foie et voies biliaires	13721	13546	0,981	0,968	0,974	-0,013	*	13644	0,989	0,983	0,986	-0,006	
02.1.06 Tumeur du Pancréas	18791	18724	0,99	0,986	0,988	-0,004		18790	0,994	0,994	0,994	0	
02.1.07 Tumeur du Larynx	1466	1465	0,945	0,945	0,945	-0,001		1474	0,968	0,973	0,971	0,005	
02.1.08 Tumeur du Poumon	48378	48254	0,984	0,981	0,982	-0,003		48422	0,99	0,991	0,99	0,001	
02.1.09 Mélanome malin de la peau	2700	2721	0,957	0,964	0,961	0,008		2727	0,972	0,982	0,977	0,01	
02.1.10 Tumeur du sein	19956	19924	0,982	0,981	0,981	-0,002		19965	0,989	0,989	0,989	0	
02.1.11 Tumeur du col de l'utérus	1259	1253	0,967	0,963	0,965	-0,005		1263	0,982	0,985	0,983	0,003	
02.1.12 Tumeur autre partie de l'utérus	4366	4291	0,976	0,96	0,968	-0,017		4327	0,986	0,977	0,982	-0,009	
02.1.13 Tumeur de l'ovaire	5281	5255	0,981	0,976	0,979	-0,005		5277	0,989	0,988	0,988	-0,001	
02.1.14 Tumeur de la prostate	14127	14063	0,979	0,975	0,977	-0,005		14121	0,987	0,987	0,987	0	
02.1.15 Tumeur du rein	5439	5372	0,972	0,96	0,966	-0,012		5422	0,981	0,978	0,98	-0,003	
02.1.16 Tumeur de la vessie	8157	8136	0,975	0,973	0,974	-0,003		8151	0,986	0,985	0,986	-0,001	
02.1.17 Tumeur cerveau et système nerveux central	6340	6186	0,98	0,956	0,968	-0,024	**	6271	0,992	0,981	0,987	-0,011	
02.1.18 Tumeur de la thyroïde	578	569	0,949	0,934	0,942	-0,016		568	0,988	0,971	0,979	-0,017	
02.1.19 Maladie de Hodgkin et lymphomes	7622	7617	0,967	0,967	0,967	-0,001		7645	0,983	0,986	0,984	0,003	
02.1.20 Leucémie	9424	9458	0,969	0,973	0,971	0,004		9465	0,983	0,987	0,985	0,004	
02.1.21 Autres Tumeur des tissus lymphoïdes et hématopoiétiques	5080	5049	0,965	0,959	0,962	-0,006		5078	0,981	0,981	0,981	0	
02.1.22 Autres Tumeurs malignes	35181	36290	0,917	0,945	0,931	0,032	****	35331	0,967	0,972	0,97	0,004	
02.2 Tumeurs non malignes	12096	12295	0,911	0,926	0,918	0,016	**	12174	0,966	0,972	0,969	0,006	
03 Maladies du sang et hématopoiétiques	4284	3974	0,849	0,788	0,817	-0,072	****	4182	0,918	0,896	0,907	-0,024	*
04.1 Diabète sucre	19071	18667	0,961	0,941	0,951	-0,021	****	18900	0,98	0,972	0,976	-0,009	
04.2 Autres maladies endocriniennes, nutritionnelles et métaboliques	17022	16785	0,926	0,913	0,919	-0,014	**	16858	0,972	0,962	0,967	-0,01	
05.1 Démences	29736	30433	0,952	0,974	0,963	0,023	****	30321	0,967	0,986	0,976	0,02	****
05.2 Abus d'alcool	3926	4002	0,906	0,924	0,915	0,019		3912	0,976	0,973	0,974	-0,004	
05.3 Pharmacodépendance, toxicomanie	386	359	0,83	0,772	0,8	-0,07	*	386	0,984	0,984	0,984	0	
05.4 Autres troubles mentaux et du comportement	6007	5949	0,902	0,893	0,897	-0,01		5943	0,966	0,956	0,961	-0,011	
06.1 Maladie de Parkinson	10654	10657	0,974	0,975	0,975	0		10666	0,984	0,985	0,985	0,001	
06.2 Maladie d'Alzheimer	29731	29721	0,983	0,983	0,983	0		29783	0,99	0,991	0,991	0,002	
06.3 Autres maladies du système nerveux et des organes des sens	18973	18821	0,926	0,918	0,922	-0,008		18922	0,973	0,97	0,971	-0,003	
07.1.1 Infarctus aigu du myocarde	21285	21523	0,957	0,967	0,962	0,011	*	21519	0,973	0,984	0,979	0,011	*
07.1.2 Autres Cardiopathies ischémiques	28969	28626	0,952	0,941	0,947	-0,012	***	28840	0,974	0,969	0,972	-0,004	
07.2 Autres maladies du cœur	80184	80437	0,953	0,956	0,954	0,003		80491	0,976	0,98	0,978	0,004	
07.3 Maladies cérébro-vasculaires	49197	49146	0,953	0,952	0,952	-0,001		49326	0,976	0,979	0,977	0,003	
07.4 Autres maladies de l'appareil circulatoire	39315	38793	0,939	0,926	0,932	-0,013	****	38939	0,977	0,968	0,972	-0,01	**
08.1 Grippe	2160	2145	0,967	0,96	0,964	-0,007		2154	0,987	0,984	0,985	-0,003	
08.2 Pneumonie	20021	20057	0,956	0,958	0,957	0,002		20080	0,977	0,98	0,978	0,003	
08.3.1 Asthme	1267	1228	0,952	0,923	0,937	-0,031		1261	0,975	0,97	0,972	-0,005	
08.3.2 Autres maladies chroniques des voies respiratoires inferieures	15859	15999	0,951	0,959	0,955	0,009		15975	0,976	0,983	0,98	0,007	
08.4 Autres maladies de l'appareil respiratoire	25548	25642	0,931	0,935	0,933	0,004		25546	0,971	0,971	0,971	0	
09.1 Ulcère gastro duodénal	1334	1358	0,915	0,931	0,923	0,018		1347	0,963	0,972	0,968	0,01	
09.2 Cirrhoses, fibroses et hépatites chroniques	10736	10754	0,957	0,958	0,958	0,002		10783	0,978	0,982	0,98	0,004	
09.3 Autres maladies de l'appareil digestif	27066	27077	0,927	0,927	0,927	0		27088	0,971	0,972	0,972	0,001	
10 Maladies de la peau et du tissu sous cutané	2561	2674	0,847	0,884	0,865	0,044	***	2548	0,964	0,959	0,961	-0,005	
11.1 Arthrite rhumatoïde et ostéoarthrite	888	841	0,908	0,86	0,884	-0,053	*	860	0,977	0,946	0,961	-0,032	
11.2 Autres maladies du système ostéoarticulaire et des muscles	5577	5541	0,815	0,81	0,813	-0,006		5437	0,96	0,936	0,948	-0,025	**
12.1 Maladies du rein et de l'urètre	12850	12503	0,935	0,91	0,922	-0,027	****	12627	0,975	0,958	0,966	-0,017	***
12.2 Autres maladies génito-urinaires	5108	5090	0,895	0,892	0,894	-0,004		5112	0,959	0,959	0,959	0,001	
13 Complications de grossesse et d'accouchement	66	19	0,947	0,273	0,424	-0,712	****	66	1	1	1	0	
14 Certains affections dont l'origine est périnatale	2374	2422	0,94	0,959	0,949	0,02		2377	0,998	0,999	0,999	0,001	
15 Malformations congénitales et anomalies chromosomiques	2489	2147	0,923	0,796	0,855	-0,137	****	2395	0,992	0,955	0,973	-0,038	**
16.1 Syndrome de mort subite du nourrisson	211	215	0,93	0,948	0,939	0,019		212	0,995	1	0,998	0,005	
16.2 Causes inconnues ou non précisées	25644	25575	0,955	0,953	0,954	-0,003		25456	0,987	0,98	0,984	-0,007	
16.3 Autres symptômes et états morbides non définis	49850	50407	0,969	0,979	0,974	0,011	***	50182	0,985	0,991	0,988	0,007	*
17.1.1 Accidents de transport	4328	4205	0,964	0,936	0,95	-0,028	**	4323	0,987	0,986	0,987	-0,001	
17.1.2 Chutes accidentelles	13942	14235	0,933	0,953	0,943	0,021	***	13979	0,98	0,982	0,981	0,003	
17.1.3 Noyade et submersion accidentelle	1291	1318	0,936	0,956	0,946	0,021		1298	0,982	0,988	0,985	0,005	
17.1.4 Intoxications accidentelles	2531	2748	0,767	0,833	0,799	0,086	****	2515	0,958	0,952	0,955	-0,006	
17.1.5 Autres accidents	22325	22302	0,896	0,895	0,895	-0,001		22011	0,972	0,958	0,965	-0,014	***
17.2 Suicide	13486	13447	0,971	0,968	0,97	-0,003		13539	0,988	0,992	0,99	0,004	
17.3 Homicides	648	614	0,897	0,85	0,873	-0,052	*	639	0,992	0,978	0,985	-0,014	
17.4 Evènement dont l'intention n'est pas déterminée	2287	2302	0,673	0,678	0,676	0,007		2289	0,901	0,902	0,901	0,001	
17.5 Autre cause externe	2142	2250	0,444	0,466	0,455	0,05	***	2006	0,865	0,81	0,837	-0,063	****
18 COVID	42021	42473	0,98	0,991	0,985	0,011	***	42323	0,988	0,995	0,991	0,007	*

Note : les degrés de significativité des écarts entre effectifs prédits et réels proviennent de tests d'égalité supposant que les effectifs réels suivent des lois de Poisson - * pval < 0.2 , ** pval < 0.1, *** pval < 0.05, **** pval < 0.01 : voir https://www.cepidc.inserm.fr/documentation/tests-statistiques-relatifs-aux-indicateurs-de-mortalite-en-population

Source : Inserm – CépiDc.

Données : Base de données de test et base 2023.

Lecture : Dans le cadre de l'évaluation globale d'une campagne de production à trois modes contre une campagne « classique », sur un total de 35 181 certificats avec une cause initiale « Autres Tumeurs malignes » à identifier, le codage automatique, et à défaut, le surmodèle en identifie 36 290, soit une précision de 91,7% (sur 100 certificats que le codage automatique/surmodèle prédit « Autres Tumeurs malignes », 91,7 sont effectivement des certificats avec cause initiale de décès « Autres Tumeurs malignes ») et un rappel de 94,5% (sur 100 certificats appartenant réellement à la cause initiale de décès « Autres Tumeurs malignes », le codage automatique/surmodèle en prédit correctement 94,5). La F-measure qui combine ces deux proportions, est de 0,931. 3,2% des prédictions de cause initiale « Autres Tumeurs malignes » sont de trop par rapport aux effectifs réels (c'est-à-dire codés manuellement). Cet écart d'effectifs est significatif, c'est-à-dire que l'écart entre effectifs prédits et effectifs réels a très peu de chance d'être dû au hasard, le codage automatique/surmodèle surévaluent en moyenne la quantité de certificats « Autres Tumeurs malignes ». La reprise manuelle, qui correspond à une stratégie de ciblage des certificats les plus complexes pour les rediriger vers le codage manuel, permet d'augmenter la F-measure à 96,7% et de faire disparaître la significativité de l'écart : l'écart entre les effectifs réels et les effectifs prédits par le codage automatique/surmodèle/reprise-manuelle est trop faible pour être certain qu'il n'est pas dû au hasard.

Au total, et sur l'ensemble de la population test de référence, la combinaison du système expert, du sur-modèle et de la campagne de reprise ciblée permet d'atteindre des niveaux de cohérence avec une campagne de codage classique très élevés pour la plupart des catégories, avec une moyenne des F-mesures par catégorie à 0,974 (soit 0,06 de plus que la campagne 2022). La seule catégorie pour

laquelle la F-mesure est inférieure à 0,9 est « 17.5 Autres causes externes »[8] (comme pour la campagne 2022), ce qui invite à considérer les codages dans cette catégorie avec prudence.

La F-mesure reste inférieure à 0,95 pour 5 catégories (contre 8 pour la campagne 2022) :

- « 01.3 Hépatites virales »
- « 03 Maladies du sang et hématopoïétiques »
- « 11.2 Autres maladies du système ostéoarticulaire et des muscles »
- « 17.4 Evènement dont l'intention n'est pas déterminée »
- « 17.5 Autre cause externe »

Des écarts conséquents et significatifs ($p < 0,05$ – au moins ***) en termes d'effectifs (test de Poisson), par catégorie, se retrouvent pour les catégories suivantes :

- « 05.1 Démences » - surestimation de 2 % de l'effectif attendu (2,3 % en 2022).
- « 12.1 Maladies du rein et de l'urètre » - sous-estimation de 1,7 % de l'effectif attendu (contre 1,6 % non significatif en 2022).
- « 17.1.5 Autres accidents » - sous-estimation de 1,4 % (contre 2 % en 2022)
- « 17.5 Autre cause externe » - sous-estimation de 6,3 % (contre 10 % en 2022)

Certaines catégories dont les écarts étaient significativement différents en 2022 ne se retrouvent pas dans la campagne de 2023 : « 03 Maladies du sang et hématopoïétiques », « 07.1.1 Infarctus aigu du myocarde », « 15 Malformations congénitales et anomalies chromosomiques », « 16.3 Autres symptômes et états morbides non définis », « 17.4 Evènement dont l'intention n'est pas déterminée ».

Comme attendu, la reprise manuelle ciblée améliore la performance, et ceci largement dans les catégories ciblées que ce soit les décès sensibles (« 01.2 VIH/Sida », « 13 Complications de grossesse », « 14 », « 15 », et « 16.1 » codes périnataux, malformations congénitales et morts subites du nourrisson, représentatifs des décès des jeunes enfants) ou les catégories ciblées par la reprise IA (« 01.3 Hépatites virales », « 03 Maladies du sang », « 05.3 Pharmacologie », « 11.2 Autres maladies musculosquelettiques », « 17.1.4 Intoxications accidentelles », « 17.3 Homicides », « 17.4 Intentions indéterminées » et « 17.5 Autres causes externes ») ou encore les reprises ciblées sur les risques de tuberculose (01.1), homicides (17.3) et pharmacodépendance (05.3).

L'approche retenue de cibler la reprise manuelle sur la base des précisions par catégorie améliore aussi les rappels. Cependant, l'objectif de dépasser 97 % de précision pour chaque catégorie qui a guidé le ciblage n'est pas satisfait pour 11 catégories (34 en 2022) : en particulier 3 catégories ont des précisions inférieures à 95 % (8 en 2022) : « 03 Maladies du sang », « 17.4 Evènement dont l'intention n'est pas déterminée » et « 17.5 Autres causes externes ». Ceci peut notamment s'expliquer par des aléas de tirage et par des effets de reports entre catégories. Tout cela devra être mis en regard *via* une analyse, non réalisée aujourd'hui, des écart-types associés à ces estimateurs.

Le tableau 16 détaille ce même type d'indicateur au niveau du chapitre de la CIM-10.

[8] A titre de comparaison pour la campagne 2018/2019, les F-mesures étaient inférieures à 0,9 pour 10 catégories sur 71 (les hépatites virales, les maladies du sang et hématopoïétiques, pharmacologie, maladies de la peau, arthrite rhumatoïde, autres maladies musculosquelettiques, les maladies génito-urinaires, les intoxications accidentelles, les intentions indéterminées et les autres causes externes).

Tableau 16. Performances en termes de cohérence de chapitre de la CIM-10 de l'ensemble de la campagne 2023 et effectifs prédits sur l'ensemble de la population test de référence.

Chapitre CIM-10	Batch Iris + surmodèle + reprise manuelle						
	Effectifs réels	Effectifs prédits	Précision	Rappel	F-measure	pred/réels - 1	Signif.
1 Maladies infectieuses et parasitaires	17 445	17 198	0,978	0,964	0,971	-0,014	**
2 Tumeurs	265 884	266 065	0,994	0,995	0,995	0,001	
3 Maladies du sang et des organes hématopoïétiques et certains troubles du système immunitaire	4 284	4 182	0,918	0,896	0,907	-0,024	*
4 Maladies endocriniennes, nutritionnelles et métaboliques	36 093	35 758	0,978	0,969	0,973	-0,009	**
5 Troubles mentaux et du comportement	40 055	40 562	0,97	0,982	0,976	0,013	***
6 Maladies du système nerveux	59 199	59 223	0,985	0,985	0,985	0	
9 Maladies de l'appareil circulatoire	218 950	219 115	0,985	0,986	0,986	0,001	
10 Maladies de l'appareil respiratoire	64 855	65 016	0,98	0,982	0,981	0,002	
11 Maladies de l'appareil digestif	39 136	39 218	0,976	0,978	0,977	0,002	
12 Maladies de la peau et du tissu cellulaire sous-cutané	2 561	2 548	0,964	0,959	0,961	-0,005	
13 Maladies du système ostéo-articulaire, des muscles et du tissu conjonctif	6 465	6 297	0,964	0,939	0,951	-0,026	***
14 Maladies de l'appareil génito-urinaire	17 958	17 739	0,971	0,96	0,965	-0,012	*
15 Grossesse, accouchement et puerpéralité	66	66	1	1	1	0	
16 Certaines affections dont l'origine se situe dans la période périnatale	2 374	2 377	0,998	0,999	0,999	0,001	
17 Malformations congénitales et anomalies chromosomiques	2 489	2 395	0,992	0,955	0,973	-0,038	**
18 Symptômes, signes et résultats anormaux d'examens cliniques et de laboratoire, non classés ailleurs	75 704	75 850	0,99	0,992	0,991	0,002	
20 Causes externes de morbidité et de mortalité	62 980	62 599	0,979	0,973	0,976	-0,006	*
21 Covid	42 022	42 323	0,988	0,995	0,991	0,007	*

Note: les degrés de significativité des écarts entre effectifs prédits et réels proviennent de tests d'égalité supposant que les effectifs réels suivent des lois de Poisson, * pval<.2, ** pval<.1, *** pval<.05, **** pval<.01

Source : Inserm – CépiDc.

Champ : Population test de référence.

Lecture : A l'issue d'une simulation complète de la campagne 2023, une campagne à deux modes auraient donné 40 055 décès (effectifs réels) pour les troubles mentaux et du comportement, contre 40 562 (effectifs prédits) pour une campagne à trois modes avec IA. Cette différence est significative, c'est-à-dire que la campagne à 3 modes de 2023 code un surplus significatif de troubles mentaux par rapport à une campagne à deux modes.

4.4 Apport des étapes de reprise ciblée sur la performance globale

La reprise manuelle apporte au total un gain de performance de 9,9 points de cohérence (Code CIM-10) sur la population du test qui aurait été codée manuellement dans une campagne classique, entre une campagne entièrement codée en IA et une campagne avec l'ensemble des étapes de la stratégie manuelle (soit 1,6 point de plus que 2022). Pour autant, les apports en performance de chaque étape ne sont pas les mêmes. En rapportant le différentiel *d'accuracy* à la part de reprise manuelle, on voit que reprendre un décès sensible est 2 fois plus efficace que la reprise d'un certificat tiré aléatoirement, et reprendre un certificat ciblé par l'IA 2,1 fois plus efficace. Ceci peut renseigner sur les proportions de reprise à allouer à ces différentes étapes, sans omettre cependant l'apport de ces reprises manuelles sur la base d'entraînement et tout en prenant en compte l'intérêt en termes de santé publique.

Tableau 17. Evaluation sur la population test des gains en cohérence (accuracy) des différentes étapes de reprises manuelles telles qu'elles ont été menées pour la campagne 2023

	Test qui aurait été codés manuellement			**Ensemble de la population Test de référence**		
	Effectifs concernés	Code CIM-10	*Shortlist* européenne	Effectifs concernés	Code CIM-10	*Shortlist* européenne
Surmodèle	246 509	0,817	0,878	787 484	0,925	0,950
+ décès sensibles (manuel)	4931	0,822	0,880	6 134	0,927	0,951
+ EDP et échantillons aléatoire	77 830	0,858	0,905	102 844	0,942	0,961
+ échantillons ciblés IA CI	47 705	0,903	0,940	47 705	0,960	0,975
+ échantillons ciblés IA CA	13 831	0,915	0,948	13 831	0,965	0,979
+ dernières vérifications IA	648	0,916	0,948	651	0,966	0,979
Total	**391 454**	**0,916**	**0,948**	**958 679**	**0,966**	**0,979**

Note **:** pour sommer les effectifs reportés, un certificat est compté uniquement dans la première catégorie (ligne) où il apparait.

Source **:** Inserm – CépiDc.

Champ : Population test de référence.

Lecture : Sur la population test qui aurait été codée manuellement dans une campagne traditionnelle, pour 246 509 certificats le codage provient de uniquement de l'IA. Si les 391 454 certificats à coder manuellement l'avaient été par prédiction d'IA (sans reprise manuelle) le code CIM-10 obtenu pour la CI aurait été le même qu'en manuel dans 81,7% des cas. Coder manuellement les
4 931 décès sensibles permet d'atteindre 82,2% de cohérence (*accuracy*).

5 Evolutions de codage

5.1 Nouveautés relatives aux recommandations de l'OMS

Aucune évolution notable n'est à signaler sur les règles de codage 2023 par rapport à 2022. En effet, depuis l'adoption de la CIM-11, la CIM-10 n'est plus maintenue ou discutée au sein des différentes groupes d'experts

5.2 Nouveautés internes au CépiDc

Des implémentations pratiques des règles de la CIM-10 peuvent être spécifiques à la France. Elles ne constituent pas des ajouts à proprement parler dans la CIM-10, mais plutôt des précisions ou des points d'attention sur des éléments déjà existants, sur lesquels une mise en avant particulière a été effectuée.

Cause externe indéterminée

Les certificats avec une circonstance apparente de décès indiquant une intention « indéterminée » et avec au plus une cause mentionnant un arrêt cardiorespiratoire (ACR – code CIM-10 commençant par R9) sont codés par la cause initiale « événement non spécifié, intention indéterminée » (code CIM-10 Y34). Cette règle concerne 3 certificats en 2020, 11 certificats en 2021, et 317 certificats en 2022. Cela a un impact sur le nombre de certificats codés en Y34 avec une circonstance apparente de décès indiquant une intention « indéterminée », de 293 certificats concernés en 2020, ils sont passés à 366 en 2021, 726 en 2022 à 2 100 en 2023.

Mise à jour des codes relatifs à la résistance aux médicaments antimicrobiens et antinéoplasiques

Les codes U82 à U85, qui concernent la résistance aux médicaments antimicrobiens et antinéoplasiques, ont été modifiés par l'OMS et enrichis en 2013 avant d'être intégrés dans leur forme actuelle en 2014 dans la CIM-10. Toutefois, leur utilisation est restée limitée avant 2022, marquée par une application peu systématique, et leur absence fréquente du dictionnaire de codage. À partir de l'année de codage 2022, une structuration plus rigoureuse de leur utilisation a été mise en place, accompagnée d'une meilleure intégration dans les pratiques de codage. Cet effort s'est poursuivi en 2023. Il est essentiel de noter que ces catégories ne sont jamais utilisées pour coder une cause initiale de décès. Elles sont spécifiquement conçues pour être utilisées en tant que codes de causes associées, permettant d'identifier et de documenter des résistances spécifiques en complément d'autres causes identifiées dans la chaîne causale. Cette distinction garantit une précision accrue dans le codage des résistances médicamenteuses tout en respectant les règles internationales de codification.

Codage de la tuberculose

Un rappel des règles de codage applicables à la tuberculose a été effectué afin de renforcer la cohérence des pratiques. Certaines modifications urgentes ont d'ores et déjà été intégrées au dictionnaire de codage. L'ensemble des ajustements sera pleinement effectif à partir de l'année de codage 2024, permettant ainsi une prise en compte encore plus précise des cas de tuberculose.

Accident en trottinette électrique

A partir de 2023, les accidents impliquant une trottinette électrique sont codés en V29, conformément aux consignes de codage actualisées.

5.3 Mises à jour du dictionnaire des expressions nosologiques utilisé par Iris Muse

Aucune mise à jour ou correction majeure du dictionnaire n'a été réalisée pour 2023.

6 Spécificités de la base de données individuelles

6.1 Variable sur le mode de codage utilisé pour coder ce certificat

Dans le cadre des évolutions du processus de codage impliquant de nouveaux outils, le CépiDc a ajouté dans la base de données une variable décrivant la manière dont a été codé le certificat : elle prend les modalités codage manuel, automatique par système expert Iris/Muse, ou autre méthode de codage (par IA). Les modalités relatives au codage par IA sont modifiées par rapport à 2021 et 2022(Annexe 8.9).

L'annexe 8.9 détaille le dictionnaire des codes des données mises à disposition dans la cadre du SNDS. On rappelle que pour les certificats codés avec IA le texte diffusé correspond au libellé brut présent sur le certificat sans découpage par rang (contrairement au cas des certificats codés manuellement ou par batch automatique Iris/muse, voir annexe de production sur les données 2021).

7 Références

Godet F, Costemalle V, Aubineau Y, Fouillet A, Coudin E., « Causes de décès en France en 2023 : des disparités territoriales », Drees ; Etudes et Résultats, n°1342, 2025

Fouillet A., Aubineau Y., Godet F., Costemalle V., Coudin É., « Grandes causes de mortalité en France en 2023 et tendances récentes », Bull Épidémiologique Hebdomadaire 2025 ; (13):218-43.

Hebbache, Z. *et al.* (2025), Rapport de production - Année de décès 2022 - Document de travail du CépiDc n5/2025

Hebbache, Z. *et al.* (2024), Rapport de production - Année de décès 2021 - Document de travail du CépiDc n4/2024

Hebbache, Z. *et al.* (2023), « Rapport de production – Années de décès 2018 et 2019 – Données définitives », Document de travail du CépiDc n3/2023

Coudin E. et Robert A., « Les statistiques sur les causes de décès : classer et coder dans la classification internationale des maladies », Courrier des statistiques n°12, 2024

Zambetta E. *et al.*, « Codage des causes de décès de 2018 et 2019 en CIM10 - Approche combinant *deep learning*, système expert et codage manuel ciblé », CépiDc, Document de travail n°2, 2023

Falissard L. *et al.* (2020), "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

Graves A. *et al.*, « Bidirectional LSTM Networks for Improved Phoneme, Classification and Recognition », Conference: Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005, 15th International Conference, Warsaw, Poland, September 11-15, 2005, Proceedings, Part II, 2005

Zambetta E. *et al.*, « 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. 2024;188(105462)

Vaswani A. *et al.*, « Attention is all you need », in Advances in Neural Information Processing Systems, pages 5998–6008, 2017

Papineni K. *et al.*, « *BLEU: a Method for Automatic Evaluation of Machine Translation* », Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 311-318, 2002

Eurostat, European shortlist of causes of death, 2012

World Health Organization, International Statistical Classification of Diseases and Related Health Problems, 10th revision, 2019.

World Health Organization, "The WHO application o fICD-10 to deaths during pregnancy, childbirth and puerperium: ICD MM", Juin 2012

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

8 ANNEXE

8.1 Version 2017 du volet médical du certificat de décès général (28 jours et plus)

VOLET MÉDICAL À remplir et à clore par le médecin ayant constaté le décès – *Renseignements confidentiels et anonymes*

INFORMATIONS RELATIVES AU DÉFUNT

Commune de décès : Code postal :

Commune de domicile : Code postal :

Date de décès : ☐ date réelle OU ☐ constatée

Date de naissance :

Sexe : ☐ masculin ☐ féminin

CAUSES DU DÉCÈS

PARTIE I **Maladie(s) ou affection(s) morbide(s) ayant directement provoqué le décès.**
Il s'agit de la maladie, du traumatisme, de l'intoxication, de la complication ayant entraîné la mort (et non du mécanisme de décès comme une syncope, un arrêt cardiaque…).

Intervalle entre le début du processus morbide et le décès
En heures, jours, mois ou ans

a)
due à ou consécutive à : b)
due à ou consécutive à : c)
due à ou consécutive à : d)

La dernière ligne remplie doit correspondre à la cause initiale

PARTIE II **Autres états morbides, facteurs ou états physiologiques (grossesse…) ayant contribué au décès, mais non mentionnés en Partie I**

INFORMATIONS COMPLÉMENTAIRES *(cocher la case appropriée pour chaque point)*

LIEU DU DÉCÈS
☐ Domicile (du défunt ou autre) ☐ Établissement de santé public
☐ EHPAD, maison de retraite ☐ Établissement de santé privé
☐ Voie publique ☐ Établissement pénitentiaire
☐ Autre lieu ou indéterminé

MORT SUBITE S'agit-il d'un décès brutal et inattendu, évocateur de mort subite* ?
☐ oui ☐ non ☐ ne sait pas
** décès non traumatique (adulte, enfant, nourrisson) avec mode de survenue brutal (en moins d'une heure ou probablement) et inattendu (exclusion des maladies chroniques au stade terminal)*

CIRCONSTANCES APPARENTES DU DÉCÈS
☐ Mort naturelle ☐ Faits de guerre
☐ Accident ☐ Complications de soins médicaux, chirurgicaux
☐ Suicide ☐ Investigations en cours
☐ Atteinte à la vie d'autrui ☐ Indéterminées

EN CAS DE MORT VIOLENTE (accidentelle, délictuelle, suicidaire, criminelle)
Précisez le lieu de survenue de l'événement déclencheur :
☐ Domicile ☐ Lieu de sport ☐ Voie publique
☐ Commerce ☐ Local industriel, chantier ☐ Exploitation agricole
☐ Établissement accueillant du public ☐ Autre lieu ou indéterminé

GROSSESSE La femme décédée était-elle enceinte ?
☐ non, pas au cours de l'année précédant le décès ☐ pas au moment du décès, mais grossesse terminée depuis 42 jours ou moins ☐ pas au moment du décès, mais grossesse terminée depuis plus de 42 jours et moins d'1 an
☐ oui, au moment du décès ☐ ne sait pas
La grossesse a-t-elle contribué au décès ? ☐ oui ☐ non ☐ ne sait pas

ACTIVITÉ PROFESSIONNELLE Le décès est-il survenu lors d'une activité professionnelle* ?
☐ oui ☐ non ☐ ne sait pas
** toute activité source de revenu (y compris au domicile), les trajets domicile-travail, les déplacements professionnels, etc*

RECHERCHE DE LA CAUSE DU DÉCÈS
Une recherche de la cause du décès a-t-elle été demandée ?
☐ oui, recherche médicale ☐ oui, recherche médico-légale ☐ non
Si oui, un volet médical complémentaire sera établi ultérieurement par le médecin ayant réalisé le diagnostic des causes de décès

SIGNATURE *Nom lisible et cachet obligatoire du médecin*

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

Le certificat peut être saisi électroniquement à l'adresse suivante https://sic.certdc.inserm.fr

8.2 Version 2017 du volet médical du certificat de décès néonatal (moins de 28 jours)

VOLET MÉDICAL À remplir et à clore par le médecin ayant constaté le décès – *Renseignements confidentiels et anonymes (* instructions en annexe)*

INFORMATIONS RELATIVES À L'ENFANT

Commune de décès : Code postal :

Commune de domicile : Code postal :

Date et heure de décès : à h

Date et heure de naissance* : à h

Sexe : ☐ masculin ☐ féminin ☐ indéterminé

Apgar à 1 minute : Âge gestationnel en semaines révolues d'aménorrhée : Poids de naissance en grammes :

INFORMATIONS RELATIVES À L'ACCOUCHEMENT

Naissance : 1. unique 2. gémellaire 3. triple 4. quadruple 5. quintuple

Numéro d'ordre de l'enfant si grossesse multiple :

Lieu d'accouchement : 1. établissement de santé 2. domicile 3. autre

Présentation : 1. sommet 2. autre céphalique 3. siège 4. autre

Début du travail : 1. spontané 2. déclenché 3. césarienne avant travail

Mode d'accouchement* : 1. voie basse non instrumentale 2. extraction instrumentale par voie basse 3. césarienne

Transfert ou hospitalisation particulière *de l'enfant : 1. oui 2. non

INFORMATIONS RELATIVES AUX PARENTS *(inscrire le code approprié)*

MÈRE Année de naissance :

Nationalité (en clair) :

Profession* (en clair) :

exercée pendant la grossesse : 1. oui 2. non 3. chômage 4. autre situation

État matrimonial : 1. célibataire 2. mariée 3. veuve 4. divorcée

La mère vit-elle en couple ? 1. oui 2. non

Nombre total de grossesses, y compris grossesse pour cet enfant :

Nombre total d'accouchements, y compris accouchement pour cet enfant* :

PÈRE Profession* (en clair) :

exercée pendant la grossesse : 1. oui 2. non 3. chômage 4. autre situation

CAUSES DU DÉCÈS *(*Lire les instructions de remplissage en annexe)*

CAUSE FŒTALE OU NÉONATALE* déterminante de la mort – Affection(s) morbide(s) ayant directement provoqué le décès.

Il s'agit de la maladie, du traumatisme, de l'intoxication, de la complication ayant entraîné la mort (et non du mécanisme de décès comme une syncope, un arrêt cardiaque...).

a)

due à ou consécutive à : b)

due à ou consécutive à : c)

Autre(s) cause(s) fœtale(s) ou néonatale(s) associées :

CAUSE OBSTÉTRICALE OU MATERNELLE* déterminante de la mort :

Autre(s) cause(s) obstétricale(s) ou maternelle(s) associée(s)* :

INFORMATIONS COMPLÉMENTAIRES *(cocher la case appropriée pour chaque point – *Lire les instructions de remplissage en annexe)*

LIEU DU DÉCÈS

☐ Domicile (du défunt ou autre)
☐ Établissement de santé public
☐ Voie publique
☐ Établissement de santé privé
☐ Autre lieu ou indéterminé

MORT INATTENDUE DU NOURRISSON S'agit-il d'un décès brutal et inattendu*?

☐ oui ☐ non ☐ ne sait pas

** décès non traumatique du nourrisson avec mode de survenue brutal (en moins d'une heure ou probablement) et inattendu*

CIRCONSTANCES APPARENTES DU DÉCÈS

☐ Mort naturelle ☐ Faits de guerre
☐ Accident ☐ Complications de soins médicaux, chirurgicaux
☐ Atteinte à la vie de l'enfant ☐ Investigations en cours ☐ Indéterminées

RECHERCHE DE LA CAUSE DU DÉCÈS*

Une recherche de la cause du décès a-t-elle été demandée ?

☐ oui, recherche médicale ☐ oui, recherche médico-légale ☐ non

Si oui, un volet médical complémentaire sera établi ultérieurement par le médecin ayant réalisé le diagnostic des causes de décès

SIGNATURE *Nom lisible et cachet obligatoire du médecin*

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

Le certificat peut être saisi électroniquement à l'adresse suivante https://sic.certdc.inserm.fr

8.3 Calendrier de production de l'année de décès 2023

Le calendrier ci-dessous synthétise les périodes de codage pour les décès 2023. Il affiche la période de synchronisation des données avec celles de l'Insee et la période d'entraînement et prédiction IA des données finalement utilisées. Il n'affiche pas la préparation des données pour la mettre à disposition du codage (entraînement initial des modèles, prédictions pour le ciblage du codage les moins bien codés par l'IA, tirages aléatoires…) qui sont fait en amont, ni les phases de traitements statistiques et informatiques aval conduisant à la constitution des bases. Les données individuelles 2023 ont été fournies à Eurostat le 26 juin 2025 et les tables individuelles complètes et détaillées au SNDS le 30 juillet 2025.

Nom de la tâche	Description	Début	Fin
Tirage aléatoire de 100 lots de production complets	Codage manuel	15/07/2024	20/09/2024
Echantillon démographique permanent	Codage manuel	30/07/2024	03/10/2024
Premier tirage aléatoire de 7000 rejets	Codage manuel	17/09/2024	19/12/2024
Vérifications Batch auto	Vérifications manuelles de certificats déjà codés (en batch)	12/09/2024	03/10/2024
Reprises IA - P1 - Précision CI 94%	Vérifications manuelles de certificats déjà codés (par IA)	23/09/2024	30/10/2024
Reprises IA - P2 - Précision CI 95%	Vérifications manuelles de certificats déjà codés (par IA)	24/10/2024	02/12/2024
Vérifications codage manuel ECH001 et ECH002	Vérifications manuelles de certificats déjà codés (en manuel)	21/10/2024	31/10/2024
Décès sensibles identifiés non codés par le batch	Codage manuel spécifique	06/11/2024	18/03/2025
Reprises IA - P3 - Précision CI 96%	Vérifications manuelles de certificats déjà codés (par IA)	20/11/2024	08/01/2025
Reprises IA - P4 - Précision CI 96.5%	Vérifications manuelles de certificats déjà codés (par IA)	30/12/2024	30/01/2025

Reprises IA - P5 - Précision CI 97%	Vérifications manuelles de certificats déjà codés (par IA)	15/01/2025	21/02/2025
Second tirage aléatoire de 5000 rejets	Codage manuel	20/01/2025	21/02/2025
Vérifications codage manuel ECH003, ECH005, ECH006 et ECH009	Vérifications manuelles de certificats déjà codés (en manuel ou par IA)	14/01/2025	30/01/2025
Reprises CA - P1	Vérifications manuelles de certificats déjà codés (par IA)	12/02/2025	05/03/2025
Reprises CA - P2	Vérifications manuelles de certificats déjà codés (par IA)	19/02/2025	13/03/2025
Reprises CA - P3	Vérifications manuelles de certificats déjà codés (par IA)	27/02/2025	01/04/2025
Reprises CA - P4	Vérifications manuelles de certificats déjà codés (par IA)	06/03/2025	18/03/2025
Vérifications codage manuel ECH010, ECH011, ECH012 et ECH014	Vérifications manuelles de certificats déjà codés (en manuel)	06/03/2025	13/03/2025
Vérifications codage manuel ECH015, ECH016, ECH019 et Reliquat Batch auto	Vérifications manuelles de certificats déjà codés (en manuel)	24/03/2025	28/03/2025
Vérifications des décès codés par l'IA	Vérifications manuelles de certificats déjà codés (par IA)	24/04/2025	15/05/2025
Reprises IA - Pharmaco - Tuberculose	Codage manuel spécifique	29/04/2025	16/05/2025
GOLD		14/05/2025	23/05/2025
Synchro Insee		Décembre 2024	14/15 décembre 2024
Diffusion des données causes initiales à Eurostat			26 juin 2025
Alimentation du SNDS			30 juillet 2025
Prédictions IA finales		avril 2025	

8.4 Méthode de ciblage des certificats à coder manuellement sur la base de prédiction IA

8.4.1 Ciblage des certificats sur la qualité des prédictions de la cause initiale

Le ciblage des échantillons d'IA de 2023 s'appuie sur le même modèle OLS estimé pour les ciblages réalisés depuis 2020 sans qu'une nouvelle estimation n'ait été effectuée en 2023.

On mobilise 50 000 observations codées manuellement issues de la base de test concernant les années 2020 et 2021 pour réaliser cette estimation. Les variables explicatives entrant dans le modèle sont les suivantes :

- Keras_iris_ci86 : La cause initiale prédite après passage par Iris/muse regroupée par item de la shortlist Eurostat (de loin la plus explicative), éventuellement regroupé pour les cas très rares dont les codes comportant 4 caractères et ayant une fréquence inférieure à 10 sont regroupés dans une catégorie « autres » (afin d'éviter les erreurs lors de la prédiction, puisque le modèle ne peut prédire des codes absents dans sa base d'entrainement).
- nb_mots et nb_codes : Des proxies de la longueur et de la complexité du texte du certificat (nombre de mots dans le certificat, polynôme jusqu'à l'ordre 3, nombre de codes dans la séquence),
- status_final :La capacité ou pas d'Iris/muse à aboutir à une cause initiale,
- Homog : L'homogénéité des codes proposés par le modèle de deep learning et Iris/muse,
- deux scores prédits par l'algorithme de deep learning
    - ci_prob : la probabilité associée au code de la cause initiale prédite par le modèle
    - ci_diff : la différence entre cette probabilité et la probabilité de la deuxième cause initiale la plus probable selon l'algorithme. Cette dernière variable est censée capter le pouvoir discriminant de l'algorithme dans sa prédiction.
- sexe et tranche_age : Le sexe et le groupe d'âge sont aussi inclus dans le modèle.
- Certif : le type de certificat électronique ou papier

1_ {CI prédite = CI codée} || nb_mots + nb_codes + status_final_lab + tranche_age + sexe + certif + ci_prob + ci_diff + Keras_iris_ci86 + Homog + nb_mots2 + nb_mots3 + nb_codes2 + nb_codes3

Le R2 ajusté du modèle est de l'ordre de 20 %, la précision (accuracy) de l'ordre de 80 % sur les données de test et d'entraînement.

À partir des données du test (environ 100 000 observations issues uniquement de 2020 et 2021 et codées manuellement), nous simulons l'impact d'une reprise manuelle ciblée. Pour cela, nous calculons, pour différentes proportions α allant de 0 % à 100 % de reprise manuelle, l'effet d'une correction priorisé sur les scores de confiance les plus faibles. Pour chaque valeur de α (proportion reprise manuelle), nous estimons la précision atteinte si α proportion était codé manuellement dans chaque catégorie.

Pour calculer la proportion de certificats à reprendre manuellement catégorie prédite par catégorie prédite, on se concentre sur les seules catégories prédites regroupées au niveau de la shortlist Eurostat pour lesquelles le niveau de cohérence (précision) entre prédiction par deep learning et codage manuel

est inférieur à 97 %. On calcule ainsi pour chacune des catégories de la shortlist Eurostat sur la base du test, un volume à reprendre pour dépasser une précision totale de 94 %, 95 %, 96 %, et 97 %, tout en tenant compte du fait que les certificats codés automatiquement et des certificats déjà codés sont supposés alors être correctement codés.

Ainsi, on définit :

- $eff\ codes$: le nombre de certificats déjà codés avec une cause initiale dans la catégorie (ce codage étant obtenu automatiquement ou bien codé manuellement)

- $eff_noncodes$: le nombre de certificats non codés et pour lesquels l'algorithme de deep learning prédit une cause initiale dans la catégorie. C'est parmi eux que l'on cherche à cibler ceux à coder manuellement en priorité
- eff_tot est la somme des deux
- Et le taux de codage dans la catégorie $a = \frac{eff\ codes}{eff\ tot}$ La précision totale est alors
- $P_t = (1-a)P_{ia} + a$ où P_{ia} est la précision pour les non-codés de la catégorie
- Si on fixe un seuil minimal P_t^* sur la précision totale, cela conduit à devoir atteindre une précision de P_{ia}^* sur les non-codés telle que

$$P_{ia}^* = \frac{(P_t^* - a)}{1-a} = \frac{P_t^* - \frac{eff\ codes}{eff\ tot}}{\frac{eff\ non\ codes}{eff\ tot}} = \frac{P_t^* * eff\ tot - eff\ codes}{eff\ non\ codes}$$

Or la précision pour les non-codés peut se voir comme une fonction du taux de reprise manuelle supplémentaire dans la catégorie. Elle vaut la précision simulée dans la catégorie s'il n'y a pas de reprise supplémentaire (estimée sur le test) et va jusqu'à 1 si on considère que toute la catégorie est reprise manuellement. L'inversion de cette fonction pour la valeur Pia* donne le taux de reprise manuelle à réaliser sur la catégorie, en se concentrant sur les certificats pour lesquels l'indicateur de confiance est le plus faible :

Taux reprise $comp\ cat = P_{ia}^{-1} * P_{ia}^*$ à affecter à l'effectif des non-codés prédits dans la catégorie pour avoir les effectifs à reprendre par catégorie.

$$Eff\ a\ reprendre = P_{ia}^{-1} * Eff\ non\ code$$

Tableau 18 Effectifs à reprendre manuellement pour atteindre XX % de précision dans la catégorie

	Niveau de précision cible				
	94 %	**95 %**	**96 %**	**96,5 %**	**97 %**
01.1- Tuberculose	19	2	33	0	11
01.2- SIDA (maladie VIH)	119	35	2	10	2
01.3- Hépatites virales	40	12	16	0	6
01.4- Autres maladies infectieuses et parasitaires	1 355	250	452	251	351
02.1.01-TM lèvres, bouche, pharynx	0	0	0	0	14
02.1.02-TM de l'œsophage	0	0	0	0	0

02.1.03-TM de l'estomac	0	0	0	0	0
02.1.04-TM Colon, rectum, anus	0	0	0	0	0
02.1.05-TM du foie et voies biliaires	0	0	0	0	0
02.1.06-TM du Pancréas	0	0	0	0	0
02.1.07-TM du larynx	0	0	6	14	6
02.1.08-TM du Poumon	0	0	0	0	0
02.1.09- Mélanome malin de la peau	0	0	0	0	0
02.1.10-TM du sein	0	0	0	0	0
02.1.11-TM du col de l'utérus	0	0	0	0	2
02.1.12-TM autre partie de l'utérus	0	0	0	0	0
02.1.13-TM de l'ovaire	0	0	0	0	0
02.1.14-TM de la prostate	0	0	0	0	0
02.1.15-TM du rein	0	0	0	0	0
02.1.16-TM de la vessie	0	0	0	0	0
02.1.17-TM cerveau et système nerveux central	0	0	0	0	0
02.1.18 - TM de la thyroïde	0	0	0	3	11
02.1.19 - Maladie de Hodgkin et lymphomes	0	0	0	0	48
02.1.20- Leucémie	0	0	0	0	0
02.1.21-Autres TM des tissus lymphoïdes et hématopoïétiques	0	0	0	0	0
02.1.22 - Autres Tumeurs malignes	88	351	702	702	527
02.2 - Tumeurs non malignes	85	170	198	85	283
03 - Maladies du sang et hématopoïétiques	654	96	80	80	48
04.1 - Diabète sucre	0	0	0	120	121
04.2- Autres maladies endocriniennes, nutritionnelles et métaboliques	336	211	294	211	168
05.1- Démences	0	0	0	0	40
05.2- Abus d'alcool	222	129	106	47	58
05.3 - Pharmacodépendance, toxicomanie	104	6	26	2	0
05.4 - Autres troubles mentaux et du comportement	256	98	138	59	177
06.1- Maladie de Parkinson	0	0	0	0	0
06.2 - Maladie d'Alzheimer	0	0	0	0	0
06.3 - Autres maladies du système nerveux et des organes des sens	256	307	307	205	256
07.1.1 - Infarctus aigu du myocarde	0	0	0	0	128
07.1.2 - Autres Cardiopathies ischémiques	0	0	117	177	176
07.2 - Autres maladies du cœur	0	0	263	525	657
07.3 - Maladies cérébraux vasculaires	0	0	0	503	302
07.4- Autres maladies de l'appareil circulatoire	88	527	616	351	352
08.1 - Grippe	0	25	18	12	25
08.2 - Pneumonie	0	0	0	102	102

08.3.1 - Asthme	7	9	11	7	14
08.3.2 - Autres maladies chroniques des voies respiratoires inferieures	0	0	106	107	106
08.4 - Autres maladies de l'appareil respiratoire	84	295	295	168	169
09.1 - Ulcère gastro duodénal	0	0	29	18	54
09.2 - Cirrhoses, fibroses et hépatites chroniques	0	0	45	90	46
09.3 - Autres maladies de l'appareil digestif	0	288	505	360	360
10 - Maladies de la peau et du tissu sous-cutané	247	29	38	38	104
11.1 - Arthrite rhumatoïde et ostéoarthrite	43	34	22	7	7
11.2 - Autres maladies du système ostéoarticulaire et des muscles	650	269	180	112	67
12.1-Maladies du rein et de l'urètre	183	157	210	78	105
12.2- Autres maladies génito-urinaires	0	91	120	46	90
13 - Complications de grossesse et d'accouchement	5	0	0	0	0
14 - Certains affections dont l'origine est périnatale	0	0	101	23	11
15 - Malformations congénitales et anomalies chromosomiques	37	66	74	52	81
16.1 - Syndrome de mort subite nourrisson	0	0	0	0	0
16.2 - Causes inconnues ou non précisées	636	424	1 801	1 377	1 272
16.3 - Autres symptômes et état morbides non définis	0	0	0	0	0
17.1.1 - Accidents de transport	0	0	0	0	0
17.1.2 - Chutes accidentelles	646	647	1 016	369	924
17.1.3 - Noyade et submersion accidentelle	39	7	9	3	4
17.1.4 - Intoxications accidentelles	657	55	44	21	55
17.1.5 - Autres accidents	2 540	605	846	363	484
17.2 - Suicide	0	0	0	0	0
17.3 - Homicides	237	14	11	4	0
17.4 - Evènement dont l'intention n'est pas déterminée	477	49	42	21	14
17.5 - Autre cause externe	850	20	31	11	20
18 – Covid	0	0	0	0	0
Total	10 960	5 278	8 910	6 734	7 858

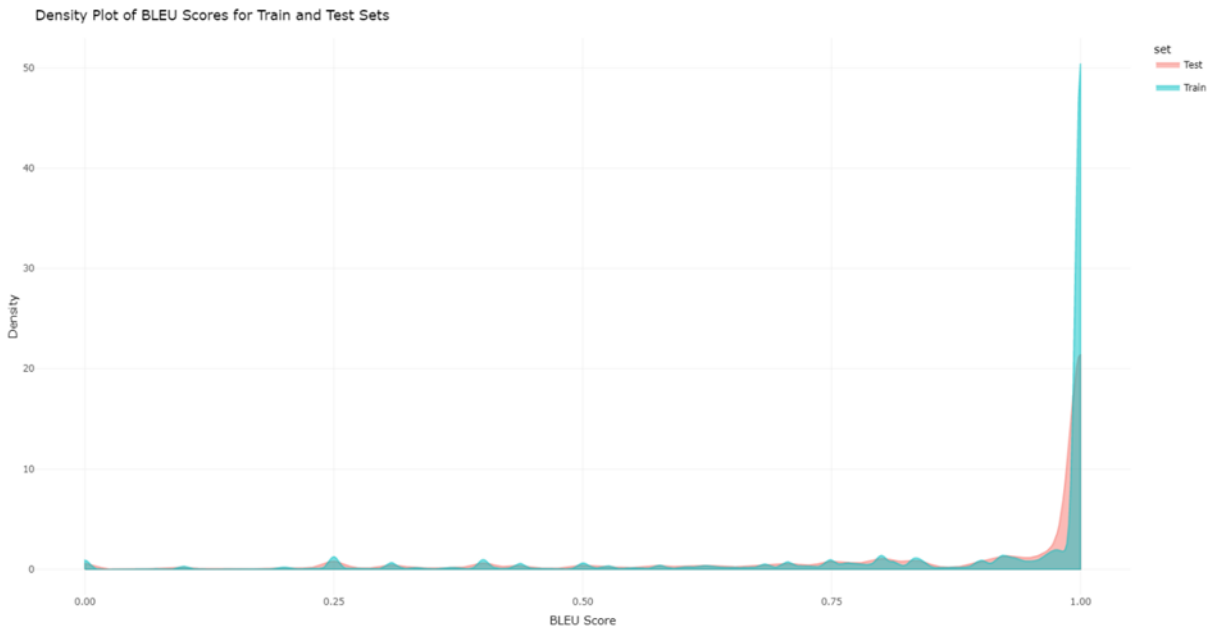
8.4.2 Ciblage des certificats sur la qualité des prédictions des causes associées

Le ciblage des certificats de décès repose ici sur un modèle de prédiction du score BLEU (voir Annexe 8.10 pour la définition détaillée de cette métrique). Ce modèle permet d'identifier les sorties de l'IA présentant les scores BLEU les plus faibles, c'est-à-dire les prédictions de moindre qualité. Ainsi, l'intervention humaine peut être concentrée sur les cas critiques, là où le modèle automatique est le moins fiable.

Les données d'entraînement et de test proviennent d'un ensemble annoté manuellement. Le jeu d'entraînement comprend 506 802 observations, filtrées à partir du codage manuel utilisé pour entraîner le modèle transformers, dont 20 % servent à la validation. Le modèle est ensuite évalué sur 335 035 observations, afin de comparer les scores BLEU prédits aux scores BLEU réels.

Plusieurs modèles de prédiction du score BLEU ont été testés. L'évaluation repose sur la capacité du modèle à identifier les certificats correspondant aux x scores les plus faibles, en comparant les rangs des scores prédits et observés, plutôt que leurs valeurs absolues. En effet, ces dernières ne sont pas directement comparables, car la majorité des scores se concentrent autour de 1, tandis que les scores faibles (< 0,5) sont très rares (voir Figure 6).

Figure 6. Distribution du score BLEU réel dans les bases de Train et de Test



Cette distribution déséquilibrée a un impact direct sur la performance du modèle de régression : il devient moins sensible aux faibles scores, et tend à les surestimer. Pour mieux comprendre cette relation, une matrice de confusion est utilisée afin de visualiser la correspondance entre les ordres prédits et réels.

Lorsqu'on soumet manuellement 10 % des scores BLEU les moins bien prédits,

- une sélection aléatoire ne permet d'identifier que 10 % des certificats réellement associés aux plus faibles scores
- un modèle Tobit, capable de gérer les valeurs censurées, améliore cette identification à 28 %
- une approche par réseau de neurones porte cette performance à 46 %

Ainsi, la priorisation des certificats à l'aide d'un modèle estimant le score BLEU réel s'avère nettement plus efficace qu'un tri aléatoire, permettant un gain significatif d'efficacité dans le volume de certificats à traiter.

Le modèle de prédiction sélectionné est le suivant :

```
model2 = Sequential()
model2.add(BatchNormalization())
model2.add(Dense(units=128, input_dim=X_train2.shape[1], activation='relu'))
model2.add(BatchNormalization())
model2.add(Dense(units=64, activation='relu'))
model2.add(BatchNormalization())
model2.add(Dense(units=32, activation='relu'))
model2.add(Dropout(0.2))
model2.add(Dense(units=1, activation='relu'))
model2.compile(loss='mean_squared_error', optimizer='adam')
```

Le réseau de neurones utilisé est un modèle de régression conçu pour prédire un score de qualité compris entre 0 et 1, à partir de plusieurs variables numériques décrivant un certificat de décès.

Il est constitué de plusieurs couches denses comportant respectivement 128, 64 et 32 neurones, activées par la fonction ReLU. Des couches de normalisation de lot (batch normalization) sont intercalées afin de stabiliser l'apprentissage, tandis qu'un mécanisme de dropout est appliqué pour limiter le surapprentissage (overfitting). La dernière couche du réseau produit une valeur unique, correspondant au score de qualité prédit. L'entraînement du modèle s'effectue à l'aide de l'optimiseur Adam, en minimisant la perte MSE (Mean Squared Error), ce qui permet d'ajuster les prédictions au plus près des valeurs réelles.

Les variables explicatives utilisées combinent des informations issues du certificat de décès, et des représentations textuelles dérivées des embeddings du modèle transformers. Un embedding est une représentation numérique des mots d'un texte : chaque mot (ou token) est transformé en un vecteur de nombres qui capture son sens et ses relations sémantiques avec les autres mots. Grâce à l'apprentissage du modèle transformers, ces représentations vectorielles sont optimisées de manière à ce que les mots ayant un sens proche soient proches dans l'espace vectoriel.

Les variables explicatives utilisées dans le modèle regroupent plusieurs types d'informations :

- les caractéristiques du certificat de décès, qui décrivent le contexte général du certificat
  - le **type de certificat** (*électronique ou papier*),
  - l'**âge**, la **tranche d'âge** et le **sexe** du défunt,
  - le **lieu du décès** (*domicile, hôpital, voie publique*),
  - les **circonstances du décès**
- Probabilités associées à la séquence de causes multiples prédite, qui proviennent des sorties du modèle transformers
  - le **score de probabilité maximale** pour chaque code de cause prédit,
  - la **différence de probabilité** entre les deux codes les plus probables, indicatrice du degré d'incertitude du modèle pour chaque code de cause prédit.
- Informations textuelles et structurelles, qui décrivent la composition du certificat et la richesse de son contenu
  - le **nombre de mots** dans le texte brut du certificat,
  - le **nombre de codes CIM-10** prédits,
  - le **nombre de lignes** présentes dans le certificat.
- Représentations vectorielles (embeddings) issus du modèle transformers qui enrichissent l'entrée du modèle
  - les **embeddings du texte brut** (provenant de l'encodeur) reflètent la **qualité linguistique du texte d'entrée**, qui influence directement la qualité des causes prédictes ;

- les **embeddings des codes prédits** (issus du décodeur) capturent la **qualité** des causes associées, essentielles à la prédiction du **score BLEU**.

Une matrice de dimension (100 × 514) est ainsi générée, puis réduite à une matrice (3 × 514) à l'aide de techniques de réduction dimensionnelle telles que MaxPooling, MinPooling et MeanPooling.

Les certificats envoyés en reprise manuelle sont au nombre de 10 296, répartis en quatre lots, ordonnés selon les scores BLEU prédits les plus faibles. Ce nombre a été fixé en fonction de la capacité de codage de l'équipe. Pour la campagne de codage de 2023, seuls les 3 premiers lots de reprises ont été traités par l'équipe de codage, tandis que le 4ème a été codé automatiquement par l'IA.

8.5 Description des bases d'entrainement et des bases de test des modèles

	Train				Test		
Echantillon	Batch	Manuel	Total		Batch	Manuel	Total
ech2011_2015	0	0	2767165				0
ech2016_2017 - codage manuelle	0	298898	298898		0	187056	187056
ech2016_2017 - codage batch	670483	0	670483		11	0	11
ech2018 - codage batch	368604	0	368604				0
ech2018 - ECH001 - EDP	0	4857	4857		0	5010	5010
ech2018 - ECH002 - Deces sensibles	0	1586	1586		0	1707	1707
ech2018 - ECH003 - IA PR1	0	1543	1543		0	1573	1573
ech2018 - ECH004 - IA PR2	0	557	557		0	487	487
ech2019 - codage batch	361748	0	361748				0
ech2019 - ECH001 - EDP	0	4834	4834		0	4853	4853
ech2019 - ECH002 - Deces sensibles	0	1624	1624		0	1622	1622
ech2019 - ECH003 - IA PR1	0	1591	1591		0	1611	1611
ech2019 - ECH004 - IA PR2	0	890	890		0	877	877
ech2020 - codage manuelle	0	155337	155337		0	121461	121461
ech2020 - codage batch	281657	0	281657		84026	0	84026
ech2021 - codage batch	322077	0	322077		96	0	96
ech2021 - ECH1 aléa	43739	15356	59095		2131	12921	15052
ech2021 - ECH10 IA PR5	0	1597	1597		0	1569	1569
ech2021 - ECH2 Faux EDP	14471	4323	18794		662	4706	5368
ech2021 - ECH3 - Deces sensibles	0	965	965		0	2904	2904
ech2021 - ECH4 - aléa 5% rejets restants	0	1325	1325		0	6031	6031
ech2021 - ECH5 - EDP	0	167	167		15403	8611	24014
ech2021 - ECH6 - IA PR1	0	3191	3191		0	3158	3158
ech2021 - ECH7 - IA PR2	0	1541	1541		0	1590	1590
ech2021 - ECH8 - IA PR3	0	2411	2411		0	2464	2464
ech2021 - ECH9 - IA PR4	0	1960	1960		0	1932	1932
ech2021 - hors_ech	0	3360	3360		0	3544	3544
ech2021 - hors_ech_a_date	0	7958	7958				0
ech2022 - codage batch	351228	0	351228				0
ech2022 - ECH001 aléa	41649	11616	53265		0	14376	14376
ech2022 - ECH002 - EDP			0		18719	11115	29834
ech2022 - ECH003 - IA PR1	0	9659	9659		0	1074	1074
ech2022 - ECH004 - IA PR2	0	4266	4266		0	474	474
ech2022 - ECH005 - IA PR3	0	7055	7055		0	836	836
ech2022 - ECH007 - Deces sensibles	0	1809	1809		0	316	316
ech2022 - ECH008 - aléa rejets restants	0	955	955		0	3000	3000
ech2022 - ECH009 - IA PR4	0	4222	4222		0	519	519
ech2022 - ECH010 - IA PR5	0	2867	2867		0	690	690

ech2022 - ECH012 - aléa rejets restants				0	3737	3737
ech2023 -ECH001 aléa	43521	11104	54625	0	13573	13573
ech2023 - ECH002 - EDP	0	0	0	18361	10191	28552
Total K5_2023 Finetuning n°1	5839339			574997		
ech2023 - ECH003 - aléa rejets restants	0	1749	1749	0	5248	5248
ech2023 - ECH005 - IA PR1	0	8805	8805	0	978	978
ech2023 - ECH006 - IA PR2	0	4223	4223	0	469	469
ech2023 - ECH009 - IA PR3	0	7079	7079	0	788	788
Total K5_2023 Finetuning n°2	5862326			582480		
ech2023 - ECH008 - Deces sensibles	0	2486	2486	0	275	275
ech2023 - ECH010 - IA PR4	0	5318	5318	0	589	589
ech2023 - ECH011 - IA PR5	0	6023	6023	0	663	663
ech2023 - ECH012 - alea rejets restants	0	0	0	0	4914	4914
ech2023 - ECH018 - alea rejets restants	0	0	0	0	45	45
ech2023 - ECHrepriseCA - P1	0	1723	1723	0	749	749
ech2023 - ECHrepriseCA - P2	0	1714	1714	0	747	747
ech2023 - ECHrepriseCA - P3	0	1705	1705	0	764	764
ech2023 - ECHrepriseCA - P4	0	251	251	0	106	106
Total K5_2023 Finetuning Final n°3	5881550			591332		

	Codage IA en production finale
Campagne 2018	201419
Campagne 2019	211142
Campagne 2021	149260
Campagne 2022	159955
Campagne 2023	146186

8.6 Définitions détaillées des vérifications et évolutions par rapport à 2022

8.6.1 Evolution des vérifications

8.6.1.1 Cohérence

Réf	Description 2022	Description 2023	Changement pour 2023
Cohérences	Une première série de vérifications consiste à vérifier les cohérences entre les codes de causes initiales et le sexe ou l'âge du défunt	Une première série de vérifications consiste à vérifier les cohérences entre les codes de causes initiales et le sexe ou l'âge du défunt	Aucun changement

8.6.1.2 Décès spécifiques

Réf	Description 2022	Changement pour 2023
DS1	Les certificats avec une mention de suicide mais non retenu en cause initiale	Aucun changement
DS2	Les certificats avec une mention d'homicide mais non retenu en cause initiale	Aucun changement
DS3a	Décès d'enfants entre 0 et 27 jours (dont les enfants nés sans vie) non codés dans l'échantillon des décès sensibles	Aucun changement
DS3b	Age entre 28 jours et 6 mois inclus et mention de code P non codés dans l'échantillon des décès sensibles	Aucun changement
DS3c	Age > 6 mois et mention de code P non codés dans l'échantillons des décès sensibles	Aucun changement
DS4	Décès d'enfants entre 28 jours et 15 ans avec une mort violente ou pathologie spécifique (dont codes P non codés dans l'échantillon des décès sensibles (cf 8.6)	Aucun changement
DS5	Certificats avec mention de SIDA/VIH non codés dans l'échantillon des décès sensibles	Aucun changement
DS6	Certificats de morts maternelles non codés dans l'échantillon des décès sensibles	Aucun changement
DS10	Les certificats ayant une cause initiale de la catégorie intention indéterminée (commençant par Y1 ou Y2 ou Y3[0-4])	Aucun changement
DS11	Création 2022 : Regroupement de DS11a et DS11 b avec les changements. Abandon définitif CI U071/U072 avec cancer mentionné en Partie I (comme réalisé sur le codage IA pour les données 2021) / Abandon pour le codage manuel (seul concerné) : CI a été modifiée par un codeur au détriment de U071/U072 / Abandon Démence Alzheimer + U071/U072 avec CI U071/U072.	Aucun changement

DS11a	Abandon en 2022 au profit de DS11 : Certains certificats spécifiques avec mention de Covid-19 (cf. 8.7) (vérification de l'application de règles complexes de codage, encore récentes)	
DS11b	Abandon en 2022 au profit de CovidLong_PostCovid et Vaccination : Les certificats ayant une mention de vaccination de Covid-19, post Covid-19, Covid-19 guéri ou de Covid-19 long	
CovidLong_ PostCovid	Création en 2022 : Division de DS11. Les certificats ayant une mention de covid long ou post covid	Aucun changement
Vaccination	Création en 2022 : Division de DS11. Les certificats ayant une mention de vaccination de Covid-19	Les mentions de type "non vacciné contre la covid-19" ne sont plus incluses dans la vérification

8.6.1.3 Difficultés de codage connues liées au logiciel Iris

Réf	Description 2022	Changement pour 2023
M1	Abandon en 2022 au profit de M1a et M1b : Vérification des certificats avec un code D00-D48 en CI alors qu'il y a mention de cancer	
M1a	Création en 2022 : Vérification des certificats avec un code D00- D45 ou D48 en CI alors qu'il y a mention de cancer	Aucun changement
M1b	Création en 2022 : Vérification des certificats avec un code D46 ou D47 en CI alors qu'il y a mention de cancer	Aucun changement
M2	Cause initiale en Y avec un code T sur le certificat : dans ces cas-là, le plus souvent une autre pathologie doit être en CI	Aucun changement
M4	Code d'affection hématologique (D50-D89) en cause initiale alors qu'il existe un code C sur le certificat	Aucun changement

M5	Notion de surdosage AVK en partie 1	Aucun changement
M6	Codes en CI qui commencent par :'M60'ou 'M79' ou 'N94' ou 'N93'' ou 'N92' : Muse s'arrête sur ces codes même s'il y a d'autres codes sur les lignes suivantes. Abandon en 2022 des cas où ces codes barrières étaient sur la dernière ligne de la P1 ou en P2	Aucun changement
M7	Traitement du "ou" si mention de mort violente à côté du “ou”	Aucun changement
M8	D611 en CI et absence de cancer dans le certificat	Aucun changement
M9	Cause initiale à coder en C80.9 si C34.9 ou un C41.1 en cause associée	Aucun changement
M10	Mention de "sur" suivi de certains codes laissant à penser que le « » sur » » ne doit pas être considéré comme un lien de causalité, avec impact potentiel sur la CI	Aucun changement
M11	Anémies en cause initiale (D619, D649 et D539) avec un code informatif (R%%, Z%%, I46%,I269,I10%,I776,I50) présent en partie 1 autre qu’un code d’anémie	Aucun changement

8.6.1.4 Vérification de la bonne application des nouvelles règles de 2016

Réf	Description 2022	Changement pour 2023
NR1	Les hépatites virales ne peuvent plus être dues à d’autres pathologies : les certificats avec mention d’hépatites virales non retenues en cause initiale	Aucun changement
NR9	Les accidents (V01-X59) ou chutes W00-W19 doivent être dû à une autre cause	Aucun changement

8.6.1.5 Vérifications liées aux évolutions du dictionnaire (démarche des choix de code)

Réf	Description 2022	Changement pour 2023
CC_Hemorragie **_inracerebrale**	S'il y a une notion de causes externes dans un certificat avec mention d'hémorragie intracérébrale, le codage sera S068 et non I6199. Depuis 2021, un allégement est appliqué au codage IA : Ne doit pas mentionner de chute en Partie I (Plage de codes W00-W19), ni un traumatisme en S ou T en partie 1 ni de Y44.2 et Y44.4 en partie 1 et partie 2	Aucun changement
CC_Alcool	S'il y a une notion de cause externe dans un certificat avec mention d'alcool, on vérifie que le code F10.2 est approprié	Aucun changement
CC_IRA	On vérifie que le code J96.0 est approprié pour les certificats avec un libellé « IRA » et une notion de diabète ou certaines pathologies rénales/urinaires. Exclusion du libellé IRA anurique	Aucun changement
CC_noyade	Libellé « noyade » et une information de dépression (F30, F32 ou F33) en partie 1 si les circonstances apparentes du décès sont inconnues. On vérifie uniquement le code W74. *Dépression et noyade en P1 ou P2 sans mention de suicide. Exclusion des cas suivants également à la lecture du tableau puisque la CI n'est pas à modifier :* *Noyade + dépression MOD=4 et CI= X71* *Noyade + dépression MOD=5 et CI=Y21*	Aucun changement

CC_paraplegie	On vérifie que le code G822 est approprié pour coder le libellé « paraplégie » s'il y a aussi une mention d'AVC sur le certificat. En 2022, n'est plus réalisée sur le codage réalisé par les méthodes d'IA	Aucun changement
CC_choc_hemorragique_ hypovolemique	Création 2022 : uniquement sur les prédictions IA. S'il y a une notion d'acte diagnostique et thérapeutique invasif dans un certificat avec mention de choc hémorragique ou hypovolémique, on vérifie que le code R57.1 est approprié (potentiellement à reprendre en T179). En 2022 n'est plus réalisée sur le codage manuel	Aucun changement
CC_Rhabdomyolyse	Création 2022 : Exclusion de Rhabdomyolyse + chutes (W00 à W019) et Inclusion de Rhabdomyolyse + code T ou code S et CI différente de X	Aucun changement

8.6.2 Décèss entre 28 jours et 15 ans identifiés dans les décès sensibles ou à vérifier

Descriptif	**Code en CI ou en cause associée (Chapitre ou Intervalle de codes)**	**Exceptions (non considéré comme spécifique à vérifier)**
Maladie du système nerveux à l'exception de …	G%	G12%, G40%, G41%, G70%, G71%, G72%, G80%, G93%
Troubles mentaux et du comportement	F%	

Maladie de l'appareil respiratoire à l'exception de …	J%	J09%, J10%, J11%, J12%, J21%, J35%, J45%, J46%, J840
Maladie de l'appareil digestif à l'exception de …	K%	K35%, K65%
Maladie de la peau et du tissu cellulaire sous-cutané	L%	
Tumeur à l'exception de …	C01-C98	C222, C40%, C41%, C49%, C62%, C64%, C71%, C72%, C91%, C92%, C93%, C94%, C95%

8.6.3 Certificats avec mention de Covid-19 vérifiés

Les certificats concernés par la vérification en lien avec l'application des règles de codage sur le Covid-19 appartiennent aux catégories suivantes :

- Cause initiale non Covid hors cancer hors mort violente, avec U071/U072 présent en cause associée
    - -
- F019 + U071/U072 avec Cause initiale U071/U072
- G122 + U071/U072 avec Cause Initiale U071/U072
- G20 + U071/U072 avec Cause initiale U071/U072
- G231 + U071/U072 avec Cause Initiale U071/U072
- G232 + U071/U072 avec Cause Initiale U071/U072
- G318 + U071/U072 avec Cause initiale U071/U072
- G319 + U071/U072 avec Cause initiale U071/U072
- K703 + U071/U072 avec Cause initiale U071/U072
- K704 + U071/U072 avec Cause Initiale U071/U072U049 en Cause initiale
    - - « » ou » » + U071/U072
- Mention de ''contexte Covid-19'' codé en U071 en P1 sans mention de comorbidités liées à la Covid-19
- Mention de ''contexte Covid-19'' codé en U071 en P2 au lieu du code Z attendu

Les trois catégories suivantes ont été abandonnées depuis 2021 :

- Cause initiale U071/U072 avec cancer mentionné en Partie I
- Cause initiale a été modifiée par un codeur au détriment de U071/U072
- Démence Alzheimer + U071/U072 avec Cause initiale U071/U072.

8.7 Précisions sur les modèles Transformers

8.7.1 Entrainement / validation / test

La base de données permettant l'entrainement et l'évaluation du modèles K5 se compose des textes des certificats des années passées, mis en entrée des modèles et des codes CIM-10 auxquels ils correspondent (à la fois pour les causes multiples et la cause initiale). Dans la base, il y a deux types de textes, les textes bruts, tels qu'ils sont collectés et les textes « nettoyés » après passage d'étapes de standardisation et après traitement manuel. Les modèles sont exclusivement entrainés sur les textes bruts, les seuls disponibles pour les certificats à coder. La base d'apprentissage contient des certificats codés manuellement et des certificats codés par système expert.

Le tableau 8.5 décrit la base d'apprentissage finale (celle utilisée pour l'entrainement final et l'évaluation des modèle utilisés dans la production de l'année 2023).

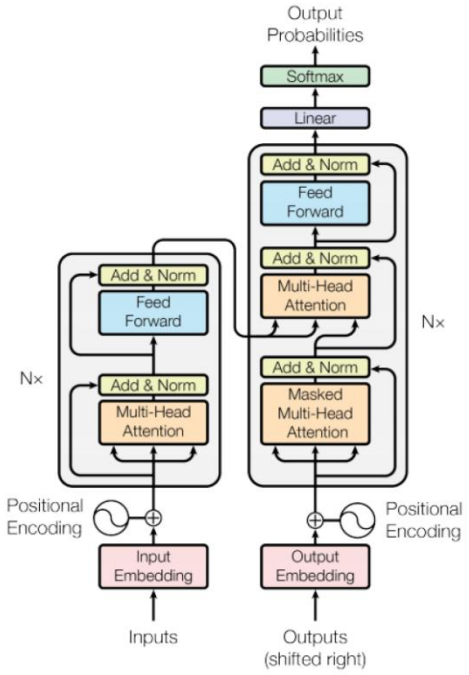
La base de test sert à évaluer la performance des modèles après entrainement. Seuls les certificats codés manuellement et provenant de tirages aléatoires sont pris en compte dans la population test de référence (391 454 certificats).

8.7.2 Description des modèles

Les modèles utilisés reprennent l'architecture Transformer propose par Vaswani *et al.* (2017). Il s'agit d'un modèle d'apprentissage supervisé « sequence-to-sequence » (seq-2-seq) avec une structure encoder/decoder (Figure 7).

La couche initiale se compose d'un plongement lexical (embedding) permettant une représentation vectorielle des mots/tokens ainsi qu'un positional encoding, qui capte l'information sur la position / le rang du mot/token dans la séquence. Dans chaque bloc d'encodage et de décodage, une couche d'attention (multiheaded attention) permet au modèle de tenir compte des relations entre les mots, même dans une longue séquence. De plus, une couche feed-forward tient compte du contexte dans son ensemble. Le décodeur contient une couche d'attention croisée, qui combine la sortie de l'encodeur avec la sortie du mécanisme d'attention à têtes multiples masquée. Le Transformer se termine par une couche softmax, dont les dimensions sont équivalentes à la taille du vocabulaire de sortie.

Figure 7. Architecture du modèle Transformers k5 implémenté pour le codage des causes de décès par AI



L'optimisation se fait grâce à l'optimizer Adam pour lequel les taux d'apprentissage (learning rates) ont été choisis selon une stratégie « warm-up » de 5 000. Les hyper-paramètres de l'optimizer Adam sont les suivants :

- beta_1 0.9
- beta_2 : 0.98
- epsilon : 1e-9

La function de perte est la sparse categorical cross-entropy.

Les autres hyperparamètres du modèle sont les suivants :

- sequence_length : 100
- batch_size : 200
- buffer_size : 5000
- embed_dim : 514
- latent_dim : 2048
- num_heads : 8
- dropout : 0.1
- epoch : 100

8.7.3 Programmes du Transformers

```
"""
## Create vocabulary: Train + Test subset
"""
tab_vocab = pd.concat([train_samples, val_samples, test])
print("Tab vocabulary :", tab_vocab.shape)

# Création du vocabulaire
inp_texts = tab_vocab['input'].to_list()
tar_texts = tab_vocab['output'].to_list()

text_vectorization_inp = Tokenizer(
    num_words=None,
    filters="-+=><!%/;.')(?°:, ",
    lower=True,
    split=' ',
)

text_vectorization_tar = Tokenizer(
    num_words=None,
    filters="-+=><!%/;.')(?°:, ",
    lower=True,
    split=' ',
)

# Input text
text_vectorization_inp.fit_on_texts(inp_texts)
voc_input = text_vectorization_inp.word_index
```

```
with open('inp_vocabulaire.json', 'w') as fp:
   json.dump(text_vectorization_inp.to_json(), fp)

pickle.dump({'config': text_vectorization_inp.get_config()}
         , open("inp_vectorization.pkl", "wb"))
# Output text
text_vectorization_tar.fit_on_texts(tar_texts)
voc_output = text_vectorization_tar.word_index

with open('tar_vocabulaire.json', 'w') as fp:
   json.dump(text_vectorization_tar.to_json(), fp)

pickle.dump({'config': text_vectorization_tar.get_config()}
         , open("tar_vectorization.pkl", "wb"))

inp_vocab_size = len(voc_input)
tar_vocab_size = len(voc_output)
print("inp_vocab_size :", inp_vocab_size) #150037
print("tar_vocab_size :", tar_vocab_size) #6333

"""
## Train and validation data
"""
inp_seq_val = text_vectorization_inp.texts_to_sequences(val_samples['input'].to_list())
inp_seq_val = pad_sequences(inp_seq_val, maxlen=sequence_length, padding="post",
truncating="post")

tar_seq_len = sequence_length + 1
tar_seq_val = text_vectorization_tar.texts_to_sequences(val_samples['output'].to_list())
tar_seq_val = pad_sequences(tar_seq_val, maxlen=tar_seq_len, padding="post", truncating="post")

val_dataset = make_dataset(buffer_size, batch_size, inp_seq_val, tar_seq_val)

inp_seq_train = text_vectorization_inp.texts_to_sequences(train_samples['input'].to_list())
inp_seq_train = pad_sequences(inp_seq_train, maxlen=sequence_length, padding="post",
truncating="post")

tar_seq_train = text_vectorization_tar.texts_to_sequences(train_samples['output'].to_list())
tar_seq_train = pad_sequences(tar_seq_train, maxlen=tar_seq_len, padding="post",
truncating="post")

train_dataset = make_dataset(buffer_size, batch_size, inp_seq_train, tar_seq_train)

"""
## Training
"""
print("Num GPUs Available: ", len(tf.config.list_physical_devices('GPU')))
print(tf.test.is_built_with_cuda())
```

```
def transformer(sequence_length,
            inp_vocab_size,
            tar_vocab_size,
            d_model,
            latent_dim,
            num_heads,
            dropout):
  encoder_inputs = keras.Input(shape=(None,), dtype="int64", name="encoder_inputs")
  x = PositionalEmbedding(sequence_length, inp_vocab_size, d_model)(encoder_inputs)
  encoder_outputs = TransformerEncoder(d_model, latent_dim, num_heads)(x)
  encoder_outputs = layers.Dropout(dropout)(encoder_outputs)
  encoder = keras.Model(encoder_inputs, encoder_outputs)

  decoder_inputs = keras.Input(shape=(None,), dtype="int64", name="decoder_inputs")
  encoded_seq_inputs = keras.Input(shape=(None, d_model), name="decoder_state_inputs")
  x = PositionalEmbedding(sequence_length, tar_vocab_size, d_model)(decoder_inputs)
  x = TransformerDecoder(d_model, latent_dim, num_heads)(x, encoded_seq_inputs)
  x = layers.Dropout(dropout)(x)
  decoder_outputs = layers.Dense(tar_vocab_size, activation="softmax")(x)
  decoder = keras.Model([decoder_inputs, encoded_seq_inputs], decoder_outputs)

  decoder_outputs = decoder([decoder_inputs, encoder_outputs])
  return keras.Model(
    [encoder_inputs, decoder_inputs], decoder_outputs, name="transformer"
  )

model = transformer(sequence_length,
              inp_vocab_size,
              tar_vocab_size,
              embed_dim,
              latent_dim,
              num_heads,
              dropout)
model.summary()

class CustomSchedule(tf.keras.optimizers.schedules.LearningRateSchedule):
  def __init__(self, d_model, warmup_steps=6000):
    super(CustomSchedule, self).__init__()

    self.d_model = d_model
    self.d_model = tf.cast(self.d_model, tf.float32)

    self.warmup_steps = warmup_steps

  def __call__(self, step):
    arg1 = tf.math.rsqrt(step)
    arg2 = step * (self.warmup_steps ** -1.5)

    return tf.math.rsqrt(self.d_model) * tf.math.minimum(arg1, arg2)
```

```
    def get_config(self):
        config = {
            'd_model': self.d_model,
            'warmup_steps': self.warmup_steps,
        }
        return config

learning_rate = CustomSchedule(embed_dim)
optimizer = tf.keras.optimizers.Adam(learning_rate,
                                     beta_1=0.9,
                                     beta_2=0.98,
                                     epsilon=1e-9)

model.compile(
    optimizer, loss="sparse_categorical_crossentropy", metrics=["accuracy"]
)

"""
## Training Model
"""
checkpoint_dir = os.path.dirname(checkpoint_filepath)

model_checkpoint_callback = tf.keras.callbacks.ModelCheckpoint(
    filepath=checkpoint_filepath,
    save_weights_only=True,
    monitor='val_loss',
    mode='min',
    save_best_only=True,
    verbose=1)

save_logs_callback = SaveLogsCallback(checkpoint_filepath + '_log.txt')

history = model.fit(train_dataset,
                    epochs=62,
                    validation_data=val_dataset,
                    callbacks=[model_checkpoint_callback, save_logs_callback])
```

8.8 Précisions sur le modèle qui sélectionne la cause initiale entre différentes propositions

Un modèle de classification en 5 classes est développé pour optimalement sélectionner la cause initiale entre la prédiction directe de cause initiale par K5, proposition la plus probable (K5top 1) mais aussi proposition la deuxième plus probable (K5top 2) et proposition la troisième plus probable (I5top 3), ainsi que l'application du système expert Iris/Muse sur la séquence de causes multiples prédite par K5 (iris5). Ces 4 propositions diffèrent de ce qui a été réalisé en 2021 et 2022 où l'on choisissait entre prédictions de deux modèles *transformers* (K5 et k4) partiellement différents (voir rapport de production 2021 et 2022). L'abandon du modèle k4 a conduit à devoir élargir le spectre de

propositions possibles à partir du seul modèle K5, d'où l'utilisation du top3 des propositions plutôt qu'uniquement la proposition la plus probable.

Ce modèle répond à un problème de classification en 5 classes, désignant parmi les types de proposition précédentes (K5top1, K5top2, K5top3, K5iris ou aucun) celle dont on retiendra la prédiction de cause initiale, et ceci en fonction des caractéristiques des certificats. L'algorithme retenu est un BiLSTM (bidirectionnel long-term short-term memory model, voir Grave et al. 2005, Bladi et al. 1999). D'autres architectures ont été testées (fnet) mais n'ont pas conduit à de meilleures performances[9].

8.8.1 Base d'apprentissage

Le modèle est entrainé sur le même jeu d'entrainement que K5, mais uniquement sur les données codées manuellement et à partir de 2016 (année à partir de laquelle on sait distinguer le mode de codage), soit 541 412 certificats, voir tableau ci-dessous. Le jeu de test est le même que celui pour évaluer K5 en 2022.

Tableau 19. Effectifs de la base d'apprentissage

	Train	**Test**
2016	149 360	93 144
2017	149 538	93 912
2018	8 543	0
2019	8 939	0
2020	155 337	121 461
2021	44 154	23 520
2022	25 541	25 186
Total	**541 412**	**357 223**

8.8.2 Problème de classification / variable cible

L'objectif du modèle de classification est de choisir le type de proposition menant à la bonne cause initiale entre K5top1 (proposition directe de cause initiale la plus probable donnée par K5), K5top2 (deuxième plus probable), K5top3 (troisième plus probable), K5Iris (proposition provenant de l'application d'Iris/muse sur la séquence de causes prédite par K5) et pas origine, indiquant qu'aucun des types de propositions ne mène à la bonne cause initiale.

[9] Accuracy sur test BilSTM .8171, Fnet .8174 (au prix d'une plus grande complexité), xgboost .789

Tableau 20. Proportion des types de proposition en base

Origine	Train : 541 412 obs	Test : 357 223 obs
iris	434 271 (80.21%)	281 259 (78.73%)
K5	42 096 (7.78%)	21 821 (6.11%)
K5_top2	24 072 (4.45%)	16 946 (4.74%)
K5_top3	8 439 (1.56%)	6 666 (1.87%)
Pas_orig	32 534 (6.01%)	30 531 (8.55%)

8.8.3 Modèle BiLSTM

On présente tout d'abord l'architecture générale du modèle utilisé. On reviendra ensuite sur ce qui compose les variables explicatives (entrées textuelles et entrées numériques).

8.8.3.1 Architecture du modèle

Il s'agit d'une architecture BiLSTM, spécifiée pour traiter séparément les entrées textuelles et les entrées numériques. La figure suivante montre l'architecture du modèle avec les différentes couches et le nombres de paramètres.

```
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================================
 input_ids (InputLayer)         [(None, None)]       0           []

 embedding (Embedding)          (None, None, 100)    10298100    ['input_ids[0][0]']

 bidirectional (Bidirectional)  (None, 256)          234496      ['embedding[0][0]']

 numerical_features (InputLayer  [(None, 19)]        0           []
 )

 concatenate (Concatenate)      (None, 275)          0           ['bidirectional[0][0]',
                                                                  'numerical_features[0][0]']

 dense (Dense)                  (None, 32)           8832        ['concatenate[0][0]']

 dropout (Dropout)              (None, 32)           0           ['dense[0][0]']

 dense_1 (Dense)                (None, 5)            165         ['dropout[0][0]']

==================================================================================================
Total params: 10,541,593
Trainable params: 10,541,593
Non-trainable params: 0
```

- ✓ **Entrée des données textuelles** : Le modèle prend en entrée une séquence de mots représentée par des identifiants numériques (`input_ids`). Cette séquence est transformée en vecteurs à l'aide d'une couche **Embedding**, qui convertit chaque mot en une représentation dense de taille 100.

- ✓ **Traitement de la séquence avec Bi-LSTM** : La séquence transformée passe ensuite dans une couche **Bidirectional LSTM**, qui analyse le texte dans les deux directions (de gauche à droite et de droite à gauche). Cela permet de mieux comprendre le contexte global de chaque mot. Cette couche LSTM utilise 256 unités.

- ✓ **Entrée des variables numériques** : En parallèle, une autre entrée est dédiée aux variables numériques (`numerical_features`), qui représentent des informations supplémentaires non textuelles comme le nombre de mots dans le certificat ou encore les probabilités de prédiction. Au total il y a 19 variables numériques ce qui explique également la taille de cette couche.

- ✓ **Fusion des informations** : Les sorties des deux parties (texte et variables numériques) sont combinées à l'aide d'une couche de **Concatenation**. Cela permet au modèle d'intégrer et de traiter les informations textuelles et numériques ensemble.

- ✓ **Régularisation** : La combinaison passe par une couche dense de 32 unités activée par une fonction **ReLU** (Rectified Linear Unit), qui aide à capturer des motifs complexes. Une couche **Dropout** avec un taux de 40% est utilisée pour ajouter de la robustesse au modèle.

- ✓ **Sortie du modèle** : Enfin, une dernière couche dense avec une activation **softmax** est utilisée pour produire la sortie. Cette couche a autant de neurones que de classes (`nb_classes`).

8.8.3.2 H-params

Batch_size = 128
Embedding dimension: 100
sequence length: 222
Vocab size: 102 973
nb_classes : 5
nb_num: 19
Dense dim: 32

Dropout: 0.4
Epoch max: 20

8.8.3.3 Stratégie d'entrainement

ModelCheckpoint (monitor='val_loss', mode='min', save_best_only=True)

Early_stopping (patience 3, val_loss)

8.8.3.4 Variables explicatives

Comme vu au-dessus, les variables textuelles et les variables numériques n'entrent pas au même moment dans le modèle. Cette distinction permet d'exploiter pleinement les variables numériques[10].

Les **variables textuelles** prises en compte correspondent aux codes CIM-10 de causes initiales proposées par les différentes classes, les causes multiples ainsi que la séquence d'entrée complète du modèle K5 (laquelle concatène les textes médicaux présents sur les certificats à des variables additionnelles d'année, sexe et âge du défunt, type et version du certificat, circonstances apparentes, etc.).

Plus précisément,

keras5_top3 : Cause initiale (CI) prédite parmi les trois meilleures par K5.
keras5_top1 : CI prédite en première position par K5.
keras5_top2 : CI prédite en deuxième position par K5.
keras5_iris : CI prédite par le modèle K5 utilisant l'approche Iris.
keras5top3_86postes : CI du top3, au niveau de la shortlist européenne (86 postes).
keras5top1_86postes : CI du top1, au niveau de la shortlist européenne (86 postes).
keras5iris_86postes : CI prédite par K5 Iris au niveau de la shortlist européenne (86 postes).
keras5top2_86postes : CI du top2, au niveau de la shortlist européenne (86 postes).
output_pred_K5 : Causes associées selon la classification CIM-10.
Status2 : Statut du codage (final ou rejet).
LieuDecesRetenu2 : Lieu du décès retenu.
inputK5 : séquence d'entrée complète du modèle K5

Les **entrées numériques** captent la complexité du certificat (nb de mots, de causes codées…), les probabilités des propositions et leurs écarts (pouvoir discriminant de K5), des indicateurs d'égalité entre le résultat de l'application d'Iris et une des propositions top3 de K5 ou encore sa présence dans les causes multiples (fiabilité de la proposition).

nb_textes : Nombre de mots présents dans le certificat.

[10] En effet, celles-ci ne passent notamment pas par la phase d'embedding dédiée au texte. Il s'agit d'une amélioration de méthode par rapport à ce qui avait été réalisé sur les données 2021 où tous les features quelle que soit leur nature numérique ou textuelle étaient concaténées en une unique séquence d'entrée.

nb_ca : Nombre de codes CIM-10 dans les causes associées.
diff_top12 : Différence entre les probabilités de la CI en top 1 et en top 2.
diff_top13 : Différence entre les probabilités de la CI en top 1 et en top 3.
diff_top23 : Différence entre les probabilités de la CI en top 2 et en top 3.
max_top1 : Probabilité de la CI en top 1.
max_top2 : Probabilité de la CI en top 2.
max_top3 : Probabilité de la CI en top 3.
nb_egal86 : Nombre de CI identiques dans les prédictions sur la shortlist Eurostat (86 catégories).
egal_top1 : Indicateur si la CI prédite en top 1 est identique à la CI Iris.
egal_top2 : Indicateur si la CI prédite en top 2 est identique à la CI Iris.
egal_top3 : Indicateur si la CI prédite en top 3 est identique à la CI Iris.
nb_egal1 : Indicateur si la CI prédite en top 1 est identique à la CI K5 Iris.
nb_egal2 : Indicateur si la CI prédite en top 2 est identique à la CI K5 Iris.
nb_egal3 : Indicateur si la CI prédite en top 3 est identique à la CI K5 Iris.
iris_in_ca : Présence de la CI Iris parmi les causes associées.
top1_in_ca : Présence de la CI prédite en top 1 parmi les causes associées.
top2_in_ca : Présence de la CI prédite en top 2 parmi les causes associées.
top3_in_ca : Présence de la CI prédite en top 3 parmi les causes associées.

8.8.4 Résultats et analyse de performance

8.8.4.1 Performance sur l'origine

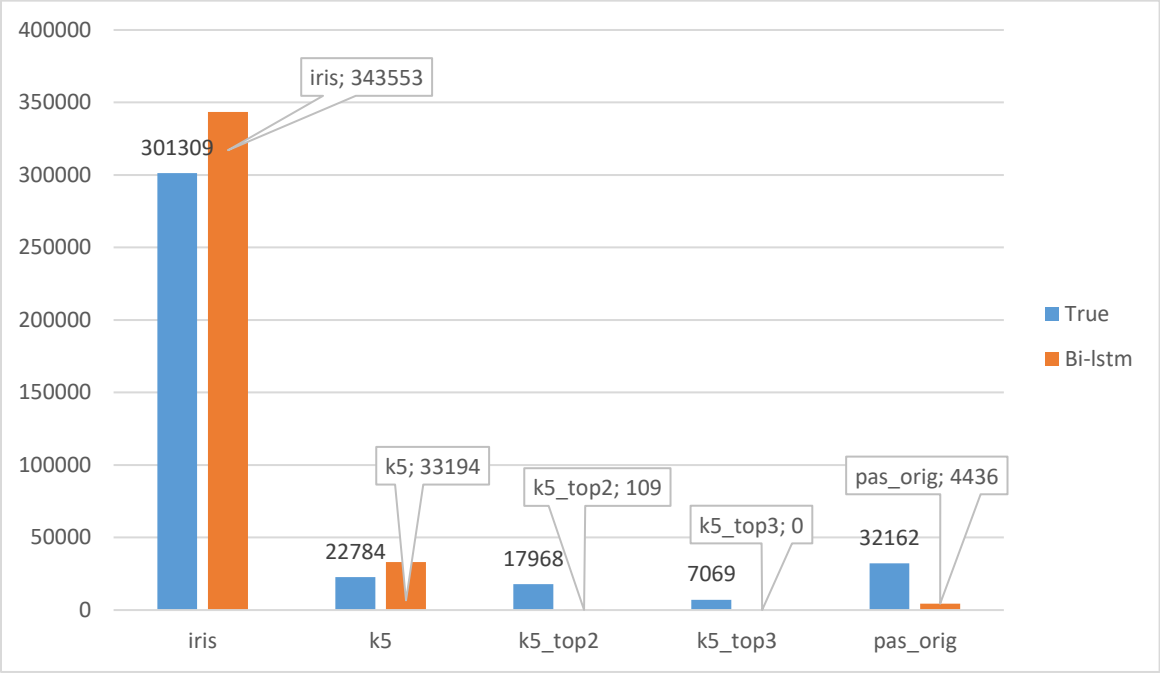
Dans 82,2% des cas (sur le test) le modèle prédit la bonne «origine», c'est-à-dire le type de proposition à retenir pour arriver à la bonne cause initiale prédite. Ce niveau est très proche de celui atteint sur la campagne 2022 (82,4%).

Train	Validation	Test
0.8559	0.8575	0.8219

Cette proportion est évidemment plus élevée sur le *train et la validation*, car le modèle a appris (a été estimé) sur ces données. L'écart de performance entre *train* et *test,* d'ailleurs, suggère une présence de léger sur-apprentissage (que l'on avait aussi un peu sur les données 2022 à surveiller.

En pratique, le surmodèle a tendance à plus souvent prédire la classe origine la plus fréquente (Iris), voire K5. Ceci est certainement lié au caractère très déséquilibré de la répartition en des classes. Notons en particulier que sur la base de test, le surmodèle n'a jamais prédit K5-top3 (il l'a fait en production cependant).

Figure 8. Répartition selon Origine - test 2023 (381 292 obs)



8.8.4.2 Performance sur la cause initiale

Au niveau du code CIM-10 de la cause initiale (c'est-à-dire le code CIM-10 de la cause initiale proposé par la classe/origine retenue), il est correct dans 81,7% des cas (sur le test) pour 81,7% en 2022, et 81,5% sur 2021.

8.9 Dictionnaire de variables dans le SNDS

A compter des décès de 2023, le bulletin de décès collecté par l'Insee a été modifié. La situation matrimoniale du défunt a été remplacée par la situation conjugale au moment du décès (marié, pacsé, en concubinage ou union libre, et dans une autre situation (veuf, divorcé célibataire). Les modalités pour l'activité professionnelle ont été également modifiées : 'inactif autre que retraité' et 'actif' sont remplacées par 'en emploi', 'au chômage' et autres.

Deux variables ont en outre été ajoutées à compter des décès 2023 : le traumatisme principal dans le cas où une cause externe est à l'origine du décès ainsi que la version de l'outil de codage utilisé pour coder le certificat de décès.

8.9.1 Table Ident

Colonne	Nom de la variable	Formats	Modalités des valeurs SNDS	Commentaires
1	Identifiant IdDécès	Varchar (64)		
2	Version du certificat	Num	1 = version de 1997 2 = version de 2017	La version 2017 apparaît à partir de 2017 (expérimentation)
3	Statut de traitement	Num	1 = Codé 2 = En cours	Les données finalisées envoyées dans le SNDS sont toutes à 1
4	Type de certificat	Num	1 = adulte 2 = néonat	Variable de mauvaise qualité ; pour identifier le type de certificat, regarder l'âge du défunt et la présence ou non de valeurs dans les variables spécifiques du certificat néonatal
5	Type de support	Num	1 = électronique 2 = papier	
6	Type de volet	Num	1 = initial 2 = Complémentaire	Les volets médicaux complémentaires apparaissent à partir de 2018

7	Département de décès	Varchar (3)		2 caractères pour métropole, 3 pour DOM TOM
8	Commune de décès	Varchar (3)		3 caractères pour métropole, 2 pour DOM TOM
9	Département de domicile	Varchar (3)		2 caractères pour métropole, 3 pour DOM TOM
10	Commune de domicile	Varchar (3)		3 caractères pour métropole, 2 pour DOM TOM
11	Date du décès	Date		
12	Lieu du décès	Num	0= Non renseigné 1 = Domicile 2 = Etablissement public de santé 3 = établissement privé de santé 4 = EHPAD, maison de retraite, 5 = voie publique 6 = autre lieu ou indéterminé 7=établissement pénitentiaire (certificat de 2017 uniquement	
13	Année de naissance	Car (4)		
14	Mois de naissance	Car (2)		
15	Sexe	Num	1 = masculin 2 = féminin 9 = indéterminé	
16	Activité professionnelle	Num	1 = Retraité 2 = Inactif autre que retraité 3 = actif 4 = en emploi 5 = au chômage 6 = autre	Variable de l'INSEE Jusqu'en 2022, modalités 1 à 3 A partir de 2023, modalités 1, et 4 à 6
17	Profession et catégorie socio-professionnelle	Varchar (2)	Nomenclature PCS INSEE	Variable de l'INSEE

18	Etat matrimonial	Num	1 = célibataire 2 = marié 3 = veuf 4 = divorcé 5 = pacsé 6 = en concubinage ou Union libre 7 = autres situations (Célibataire, veuf, divorce)	Variable de l'INSEE à partir du bulletin de l'état civil A compter des décès 2023, seules les modalités 5,6 et 7 sont utilisées
19	Cause initiale du décès	Varchar (4)	CIM-10	
20	Recherche de la cause de décès	Num	1 = non 2 =oui, résultats disponibles 3 =oui, résultats non disponibles 4 =oui recherche médicale (certificat de 2017 uniquement), 5 = oui recherche médico-légale (certificat de 2017 uniquement)	Valeur vide de 2006 à 2009, question n'était pas posée avant 2010. Valeur 2 et 3 uniquement pour les versions 1997 Valeur 4 et 5 uniquement pour les nouveaux certificats à partir de 2017
21	Grossesse "Le décès est-il survenu pendant une grossesse (ou moins d'un an après) " (ancienne version) ? La femme décédée était-elle enceinte ?" (nouvelle version)	Num	1 = non, pas au cours de l'année précédant le décès 2 = Pas au moment du décès mais grossesse terminée depuis 42 jours au moins 3 = Pas au moment du décès mais grossesse terminée depuis plus de 42 jours et moins de 1 an 4 = Oui, au moment du décès 5 = Ne sait pas	Cette variable est vide avant 2015 Libellé de la question et modalités de réponses ont changé entre la version 1997 et 2017 mais la variable reste la même
22	Délai entre fin de grossesse et décès	Varchar (4)	mois + jour, ex : 0502 = 5 mois + 2 jours après la date de fin de grossesse	Variable mal renseignée, de mauvaise qualité et uniquement sur la version 1997.

23	Ancien certif : Est-ce un Accident de travail ? Nouveau certif : Est-ce pendant une activité professionnelle ?	Num	1 = oui 2 = Non 3 = sans précision (ancien) ou Ne sait pas (nouveau)	Libellé de la question et modalités de réponses ont changé mais la variable reste la même
24	Apgar à une minute	Num	0 à 10 (score d'Apgar, 0 = mort apparente, 10 = état optimal)	Variable spécifique du certificat néonatal
25	Âge gestationnel en semaines révolues d'aménorrhée	Num	en semaines	variable spécifique du certificat néonatal
26	Poids de naissance en grammes	Num	en g	variable spécifique du certificat néonatal
27	Type de naissance	Num	1 = Unique 2 = Gémellaire 3 = Triple 4 = Quadruple 5 = Quintuple	variable spécifique du certificat néonatal
28	N° d'ordre de l'enfant si grossesse multiple	Num	1 à 5	variable spécifique du certificat néonatal
29	Lieu d'accouchement	Num	1 = Etablissement de sante 2 = Domicile 3 = Autres	variable spécifique du certificat néonatal
30	Présentation de l'enfant	Num	1 = Sommet 2 = Autres céphaliques 3 = Siège 4 = Autres	variable spécifique du certificat néonatal
31	Début du travail	Num	1 = Spontané 2 = Déclenché 3 = Césarienne	variable spécifique du certificat néonatal
32	Mode d'accouchement	Num	1 = Voie basse 2 = Extraction 3 = Césarienne	variable spécifique du certificat néonatal

33	Transfert ou hospitalisation particulière de l'enfant	Num	1 = Oui 2 = Non	variable spécifique du certificat néonatal
34	Année de naissance de la mère	Varchar (4)		variable spécifique du certificat néonatal
35	Activité professionnelle de la mère	Num	1 = En activité 2 = Non au chômage 3 = Autres 4 = Au chômage 5 = Non	variable spécifique du certificat néonatal 2006 à 2015 uniquement les 3 1ere modalités, les autres modalités arrivent avec le nouveau certificat néonatal en 2017
36	Profession de la mère exercée pendant la grossesse	Varchar (50)	libellé de la profession	variable spécifique du certificat néonatal
37	Etat matrimonial de la mère	Num	1 = célibataire 2 = mariée 3 = veuve 4 = divorcée 5 = pacsé 6 = en concubinage ou Union libre 7 = autres situations (Célibataire, veuf, divorce)	variable spécifique du certificat néonatal Mes modalités effectivement utilisées dans cette variable sont les modalités 1 à 4.
38	La mère vit elle en couple	Num	1 = Oui 2 = Non	variable spécifique du certificat néonatal
39	Nombre total de grossesses	Num		variable spécifique du certificat néonatal
40	Nombre total d'accouchements	Num		variable spécifique du certificat néonatal
41	Activité professionnelle du père		1 = En activité 2 = Non au chômage 3 = Autres 4 = Au chômage 5 = Non	variable spécifique du certificat néonatal 2006 à 2015 uniquement les 3 1ere modalités, les autres modalités arrivent avec le nouveau certificat
42	Profession du père exercée pendant la grossesse		libellé de la profession	

43	Mort subite	Num	1 = Ne sait pas 2 = Oui 3 = Non	Uniquement dans la version 2017
44	Circonstance apparente du décès	Num	1 = Indéterminé 2 = Mort naturelle 3 = Accident 4 = Suicide 5 = Atteinte à la vie du défunt 6 = Fait de guerre 7 = Investigations en cours 8 = Complications de soins 9 = Atteinte à la vie de l'enfant (certificat néonatal uniquement) ;	Uniquement dans la version 2017
45	Codification du lieu si mort violente	Num	1 = Autre lieu ou indéterminé 2 = Domicile 3 = Etablissement accueillant du public 4 = Exploitation agricole 5 = Lieu de sport 6 = Commerce 7 = Voie publique 8 = Local industriel - chantier	Uniquement dans la version 2017
46	Fiabilité de la date de décès	Num	1 = date réelle 2 = Hypothèse	Uniquement dans la version 2017
47	Mort inattendue du nourrisson	Num	1 = Ne sait pas 2 = Oui 3 = Non	variable spécifique du certificat néonatal Uniquement dans la version 2017

48	Groupe d'âge mortalité néonatale		1 = < 7 jours 2= 7 jours <= âge < 28 jours 3 = 28 jours<= âge <= 365 jours	
49	La grossesse a-t-elle contribué au décès ?	Num	1 = Oui 2 = Non 3 = Ne sait pas	Uniquement dans la version 2017
50	CauseInitialeTypeCode	Num	1 = CIM-9 2 = CIM-10 3 = CIM-11	
51	TypeCodage	Num	1=codage Iris Automatique 2= manuel *3= K5Iris* *4=K4Iris* *5=K5* *6=K4* 7=Codage IA Iris 8= Codage IA	Uniquement à partir de 2018 Décrit le mode de codage utilisé. Lee types de codage 3 à 8 impliquent des algorithmes de *deep learning (IA)* combinés ou non avec le système expert IRIS/Muse. Les modalités 7 et 8 sont disponibles depuis 2022 et les modalités 3 à 6 sont abandonnées sur 2022
52	Score de confiance	Num		Uniquement à partir de 2018 Indicateur de confiance issu de la prédiction impliquant des algorithmes de *deep learning* permettant de cibler le codage manuel. Il est compris entre 0 et 1.
53	Traumatisme principal	Varchar (4)	CIM-10	Code du traumatisme principal dans le cas où une cause externe est à l'origine de décès. Il peut être inconnu dans certains cas de causes externes et être renseigné même si une cause externe n'a pas été retenue comme cause initiale de décès.
54	Score de confiance	Varchar (50)		Indique les versions de l'outil de codage automatique (IRIS et Muse) pour le codage automatique ou manuel assisté et le modèle d'IA ainsi que le choix du surmodèle en cas de codage par IA.

8.9.2 Table des causes

Colonne	Nom de la variable	Formats	Modalités des valeurs SNDS	Commentaires
1	identifiant IdDécès	Varchar (64)		
2	N° de ligne de la cause sur le certificat	numérique	1 à 6	Numéro de la ligne de cause
3	rang de la cause	numérique	1 à 40	rang de la cause sur la ligne si plusieurs causes par ligne
4	libellé de la cause*	Varchar (400)	texte de la cause	Pour les certificats codés avec algorithme de *deep learning* (typecodage >=3) le texte de la ligne correspondra au texte brut présent sur le certificat sans découpage par rang
5	code CIM-10 de la cause**	Varchar (4)	Tables ir_cci_v* et ir_cim_v*	Ce champ va évoluer avec la CIM-11 et passera en varchar 6
6	TypeCodage			Variables non informatives à l'échelle de la ligne de cause, contenu vide.
7	Score de confiance			Variables non informatives à l'échelle de la ligne de cause, contenu vide.

* Certains libellés de causes peuvent avoir plusieurs codes CIM-10 associés. Dans ces cas-là (hors codage avec de l'IA), les codes sont indiqués à la suite (implémentation de rangs) et le texte associé n'est pas répété. C'est pourquoi on peut retrouver des codes CIM-10 sans libellé de texte associé.

** Certains libellés de causes peuvent ne pas avoir de CodeCIM correspondant : il s'agit de texte non informatif non pris en compte dans le codage. C'est pourquoi on peut retrouver des textes sans codes associés dans cette table.

8.10 Evaluation de la cohérence des causes multiples avec une campagne de codage à trois modes

L'évaluation des campagnes de codage des causes de décès a jusqu'ici mis l'accent sur la cohérence de la détermination de la cause initiale de décès. Toutefois, avant d'atteindre cet objectif, il est indispensable de transcrire et de coder en CIM-10 chacune des causes mentionnées dans les certificats de décès. Dans ce cadre, l'attention se porte désormais sur une évaluation globale de la transcription intégrale du processus morbide en codes CIM-10, dépassant ainsi la simple analyse de la cause initiale.

Pour répondre à cette exigence, un score de confiance a été élaboré afin de comparer la qualité (i.e. la cohérence) entre codage par IA, et codes réels (issus de la campagne à deux modes). Ce score, conçu pour évaluer une tâche de transformation séquentielle texte-à-texte, complète l'évaluation précédemment réalisée sur la cause initiale et renforce l'évaluation de la campagne à trois modes en y ajoutant une analyse de la traduction complète du processus morbide.

8.10.1 Score BLEU

Pour mener l''évaluation de la cohérence de la transcription des processus morbides sur les certificats de décès, le choix s'est porté sur la métrique Score BLEU, largement utilisée pour évaluer la qualité des prédictions dans des contextes de traitement automatique du langage. Ici, le BLEU Score est calculé pour chaque ligne du certificat en tenant compte des codes CIM-10 associés.

Pour obtenir une mesure à l'échelle du certificat et en particulier pour être capable de comparer les codages des processus morbides de certificats présentant des nombres de causes très différents ou des structures par ligne différente, on utilise une pondération dynamique des scores en fonction de la position des lignes et des parties du certificat. Par exemple, le score attribué à la dernière ligne de la Partie I est pondéré plus fortement, car cette ligne contient fréquemment la cause initiale de décès. À l'inverse, les lignes de la Partie II reçoivent un poids plus faible, étant donné que la cause initiale y apparaît rarement.

Le paragraphe suivant détaille la définition et la méthodologie mise en œuvre.

8.10.1.1 Définition du Score BLEU

Le score BLEU a été introduit par Papineni *et al*. en 2002 [voir 7. Références] Il permet de mesurer la précision des n-grams générés par un système de traduction automatique en les comparant à des n-grams de référence, dits « gold ».

Le calcul du score BLEU se fait en trois étapes :

1. **Précision des n-grams** : la précision pour un ordre *n* des n-grams est donnée par la formule suivante

$$Précison = \frac{Nombre\ total\ de\ ngrams\ correctement\ prédits}{Nombre\ total\ ngrams\ (gold)}$$

2. **Moyenne géométrique des précisions** : la moyenne géométrique des précisions p_n est pondérée par des poids w_n, un pour chaque ordre de n-grams. Cette moyenne est donnée par :

$$Moyenne\ géometrique\ des\ précisions = \prod_{n=1}^{N} {p_n}^{w_n}$$

Où N est l'ordre maximal des n-grams, et w_n, est le poids associé à chaque ordre de n-grams.

3. **Facteur de pénalité :** le score BLEU intègre également un facteur de pénalité BP pour corriger les biais induits par des traductions trop courtes. Ce facteur est défini comme suit :

$$BP\ (Facteur\ de\ pénalité) = \begin{cases} 1 \\ e^{(1-\frac{r}{c})} \end{cases}$$

Où *c*'est la longueur de la séquence générée et *r* est la longueur de la séquence de référence (gold).

Le score BLEU final est donné par la formule :

$$BLEU = Facteur\ de\ pénalité * Moyenne\ géometrique\ des\ précisions$$

8.10.1.2 Choix pour l'implémentation du Score BLEU

Dans cette implémentation, nous avons opté pour les paramètres suivants :

- Utilisation de n-grams de taille 2 (utilisation d'uni-grammes et de bi-grammes de mots).
- Poids égaux pour les n-grams de taille 1 et 2, soit $w_1 = w_2 = 0.5$
- Calcul du score BLEU pour chaque ligne d'un certificat, avec une pondération dynamique selon les sections du certificat détaillée par la suite.

8.10.1.3 Calcul du Score BLEU avec Pondération Dynamique

La pondération dynamique a pour objectif de rendre comparables les bleus scores de certificats ayant des tailles différentes (nombre de causes) ou des structures différentes (partie I et partie II). Elle surpondère la ligne où l'on s'attend à retrouver la cause initiale du décès et au contraire affecte un poids relatif plus faible à la partie 2.

Les lignes d'un certificat sont organisées en deux parties distinctes, chacune dotée d'une pondération spécifique. Un coefficient total de **10** est assigné à l'ensemble du certificat. On choisit une répartition pondérée et équilibrée des coefficients entre les différentes sections et lignes du certificat, tout en mettant en avant les lignes clés.

- **Première partie (Partie I)** :
  La première partie du certificat comprend jusqu'à 4 lignes. Le poids est réparti proportionnellement entre les lignes, selon leur nombre. La dernière ligne de cette partie se voit attribuer un poids **trois fois supérieur** à celui des autres lignes de la Partie I.
- **Deuxième partie (Partie II)** :

La deuxième partie, si elle est présente, reçoit une pondération qui correspond à **un tiers** de celle attribuée à la Partie I. Cette pondération réduit donc l'importance relative de cette section par rapport à la première.

Cas 1 : Pas de Partie II

Pour un certificat contenant uniquement la partie I, le poids des lignes est calculé en fonction du nombre de lignes *a* de la Partie I. L'équation pour déterminer les poids des lignes est la suivante :

$$10 = (a - 1)p1 + 3p1$$

Cela donne la solution pour (poids de la Partie I) et (poids de la dernière ligne de la Partie I) :

$$p1 = \frac{10}{a+2}; pd1 = \frac{30}{a+2}$$

Cas 2 : Avec Partie II

Dans le cas où une partie II est présente, l'équation à résoudre devient :

$$10 = (\mathrm{a} - 1)\mathrm{p1} + 3\mathrm{p1} + \frac{p1}{3}$$

Les poids sont calculés comme suit avec : poids de la Partie I, : poids de la dernière ligne de la Partie I et : poids de la Partie II

$$p1 = \frac{10}{a + \frac{7}{3}}\ ;\ pd1 = \frac{30}{a + \frac{7}{3}}\ ;\ p2 = \frac{10}{3a + 7}$$

Cela permet de pondérer les scores BLEU en fonction du nombre de lignes dans le certificat.

8.10.1.4 Exemples en pratique

Les exemples suivants illustrent les choix méthodologiques adoptés pour l'évaluation et la pondération des lignes du certificat de décès, tels qu'ils ont été décrits précédemment.

Exemple 1 :

Partie I :

Ligne 1 : traumatisme crânien

Ligne 2 : pneumopathie hypoxiémiante

Ligne 3 : arrêt cardiaque hypoxique

Partie II : chirrose child b

		Codage du processus morbide dans une campagne traditionnelle	Codage du processus morbide par algorithme d'IA par K5	Poids attribué dynamiquement	score
Partie I	Ligne 1	S069	S069		1
	Ligne 2	J189	J189		1
	Ligne 3	I469	I469		1
Partie II		K746	K746		1

Ce premier exemple illustre un codage parfait du processus morbide par l'intelligence artificielle. Dans ce cas, les poids attribués aux parties et aux lignes du certificat n'ont aucun impact sur le score final, qui atteint la valeur maximale de 1.

Exemple 2 :

Partie I :

Ligne 1 : dissociation électromécanique

Ligne 2 : hyperkaliémie

Ligne 3 : ischémie aigue membre inférieur

Ligne 4 : canule cec sur choc cardiogénique

Partie II : cardiopathie ischémique terrain vasculaire insuffisance rénale chronique

		Codage du processus morbide dans une campagne traditionnelle	**Codage du processus morbide par algorithme d'IA par K5**	**Poids attribué dynamiquement**	**score**

Partie I	Ligne 1	I500	I500	1.6	1
	Ligne 2	E875	E875	1.6	1
	Ligne 3	I743	I743	1.6	1
	Ligne 4	Y848	Z998, R570	4.7	0
Partie II		R570, I259, I99, N189	I259, I99, N189	0.5	0.6

Dans ce deuxième exemple, la transcription du processus morbide en codes CIM-10 par l'intelligence artificielle présente deux erreurs :

1. **Première erreur (ligne 4)** :
   L'algorithme a mal retranscrit le texte en code CIM-10, ce qui a entraîné un score **BLEU de 0** pour cette ligne.
2. **Deuxième erreur (Partie II)** :
   Le code « **R570 »** a été omis dans la Partie II du certificat, ce qui a donné un score **BLEU de 0,6** pour cette section.

En calculant une moyenne pondérée des scores, le certificat obtient un score global de **0,5**. L'erreur de transcription sur la ligne 4 impacte fortement ce résultat, rendant ce certificat **éligible à une reprise manuelle**. Cependant, l'algorithme a correctement codé la cause initiale, identifiée comme « **I259 »**. Cette métrique, conçue pour compléter l'analyse de cohérence liée à la cause initiale, offre une évaluation plus approfondie. En effet, elle permet d'examiner des certificats qui ne seraient pas forcément détectés comme problématiques par l'évaluation de la cause initiale seule. Cela enrichit le processus global d'analyse.

8.10.2 Population test de référence

La base de test utilisée pour évaluer la cohérence de la traduction du processus morbide est identique à celle utilisée précédemment pour analyser la cohérence de la cause initiale dans la campagne à trois modes. Cependant, certains certificats ont été exclus de la population test de référence, afin de garantir une analyse fondée sur des données cohérentes et comparables. Plus précisément, les observations où le nombre de lignes du « gold standard » ne correspond pas à celui du texte brut ont été exclues. Ces écarts proviennent d'ajouts de lignes par les codeurs pour compléter des informations manquantes (entre le texte saisi par le prestataire et celui rédigé par le médecin) ou de suppressions de lignes pour des mentions non codables, comme « cf. médecin traitant » en première ligne. Après filtrage, le nombre d'observations passe de 391 454 à 368 644 pour la part de la population test qui aurait été codée manuellement. Le nombre d'observations correspondant au codage batch reste inchangé, garantissant ainsi une base d'analyse fiable.

8.10.3 Performance sur la séquence de cause multiple

L'évaluation de la cohérence de la séquence des causes multiples se déroule en deux temps.

1. **Premier niveau de performance (cohérence) :**
   On commence par évaluer la cohérence sur la part de la population test qui aurait été codée manuellement dans une campagne traditionnelle comportant deux modes de codage.
   a. D'abord, on évalue la cohérence en utilisant uniquement le codage par IA.
   b. Ensuite, on ajoute la part des certificats repris manuellement pour vérifier la qualité de la cause initiale. (Inclut l'ensemble des reprises effectuées : échantillons aléatoires, EDP, reprises IA sur la qualité de la cause initiale, vérifications et décès sensibles.)
   c. Enfin, on inclut la part des certificats repris pour améliorer la qualité des causes multiples. On revoit en codage manuelle les scores BLEU pondérés prédits les plus faibles (voir **Annexe 8.4.2** méthode de ciblage des certificats suivant la qualité des prédictions IA sur les causes associées)
2. **Deuxième niveau de performance (cohérence) :**
   La cohérence globale est évaluée en tenant compte de la part du codage par batch dans la campagne, afin d'obtenir une évaluation complète des trois modes de codage. Ce niveau d'évaluation permet de comparer une campagne traditionnelle à deux modes de codage (batch + codage manuel) à la campagne actuelle, qui en compte trois grâce à l'introduction du codage automatique par algorithme de deep learning.

Tableau 21. Score BLEU pondéré des causes multiples prédites par le modèle de deep learning (K5), ainsi que des résultats obtenus après combinaison des prédictions du modèle avec les reprises manuelles, d'abord sur la qualité de la cause initiale, puis sur celle des causes multiples, comparées aux données qui auraient été codées manuellement dans une campagne traditionnelle.

Année	**2016**	**2017**	**2020**	**2021**	**2022**	**2023**	**Ensemble**
Transcription du processus morbide par IA	0.85	0.854	0.865	0.853	0.861	0.864	0.858
+ reprise manuelle	0.91	0.914	0.919	0.91	0.917	0.919	0.915
+ reprise manuelle sur la qualité de la séquence multiple	0.928	0.928	0.928	0.92	0.933	0.935	0.928
Nb Obs	81978	87007	118299	23008	25043	33309	368644

Dans la population test de référence, le Tableau 21 porte sur les certificats qui auraient été codés manuellement dans une campagne traditionnelle. L'algorithme de *deep learning* atteint une cohérence de 85,8 % dans la transcription du processus morbide, contre 85,6 % en 2022. En combinant le codage automatisé par IA avec le codage manuel, la performance atteint 91,5 % après les reprises manuelles, puis 92,8 % après les reprises ciblant la qualité des séquences multiples (contre 91,2 % en 2022). Cette dernière étape, introduite lors de la campagne 2023, apporte un gain supplémentaire de 1,3 point. Ces résultats mettent en évidence une bonne concordance entre les codes des causes multiples produits par l'algorithme et ceux issus d'un codage manuel, ainsi qu'une amélioration significative grâce à l'intervention humaine. Par ailleurs, la stabilité des performances de l'algorithme d'une année sur l'autre renforce la fiabilité du système dans le cadre des campagnes de codage (voir Tableau 21).

Tableau 22. Score BLEU pondéré des causes multiples obtenu en combinant les deux modes de codage *deep learning* (K5) et batch, ainsi que sur l'ensemble de la campagne de codage intégrant les trois modes.

Année	**Transcription du processus morbide par IA + batch**	**+ reprise manuelle**	**+ reprise manuelle sur la qualité de la séquence multiple**	**Nb Obs**
Ensemble	0.944	0.967	0.972	935869
2016	0.941	0.965	0.972	210641
2017	0.942	0.966	0.972	219951
2020	0.943	0.966	0.97	282622
2021	0.946	0.967	0.971	62546
2022	0.949	0.969	0.975	67531
2023	0.951	0.971	0.977	92578

Sur l'ensemble de la population test (Tableau 22), y compris les certificats codés par batch, la cohérence atteint 97,2 % pour la transcription du processus morbide, contre 96,5 % en 2022, en comparaison avec une campagne classique à deux modes de codage.

Les Tableaux 23 et 24 présentent l'évaluation de la cohérence de la transcription du processus morbide au sein de la *shortlist* Eurostat. Cette cohérence est d'abord mesurée sur la proportion de cas qui auraient été codés manuellement lors d'une campagne classique sur la population test de référence, puis sur l'ensemble de la campagne.

Elle est évaluée en moyennant les scores BLEU selon le chapitre Eurostat correspondant à la cause initiale réelle, ce qui peut être interprété comme une mesure de rappel. Deux moyennes sont calculées :

- la première (notée **BS** dans les Tableaux 23 et 24) correspond au score BLEU défini dans la section 8.10.1 de cette annexe ;
- la seconde (notée **BS1**) est calculée uniquement lorsque le score BLEU est égal à 1.

Cette distinction est introduite pour tenir compte de notre définition subjective de l'évaluation, qui repose sur des impératifs pratiques liés à la production des données. En ajoutant la moyenne BS1, nous identifions les cas de transcription parfaite par l'algorithme de *deep learning*, ce qui permet d'évaluer objectivement la séquence des causes associées tout en écartant les choix pratiques liés au processus de production.

L'algorithme de deep learning reproduit exactement la même séquence de causes que celle codée manuellement dans 56,4 % des cas (voir Tableau 23, colonne « BS1 »). Ce taux atteint 74,9 % après les reprises manuelles ciblées, puis 90,1 % lorsque le codage en *batch* est inclus (voir Tableau 24, colonne « BS1 »), contre 88,9 % en 2022.

L'introduction d'une reprise manuelle ciblée axée spécifiquement sur la qualité des causes consiste à renvoyer en codage manuel les certificats pour lesquels le score BLEU prédit est le plus faible. Cette approche permet un gain de performance notable dans plusieurs catégories (écart entre les scores BS1 avec reprise manuelle et BS1 après ajout de la reprise ciblée sur les causes multiples) cf. Tableau 23 :

- +11 points pour les Accidents de transport,
- +10 points pour les Suicides,
- +7 points pour les Homicides,
- +4 points pour les Tumeurs malignes des lèvres, de la bouche et du pharynx, les Événements dont l'intention n'est pas déterminée et les Leucémies.

De nombreuses autres catégories présentent également un gain de 2 à 3 points, confirmant que l'apport de cette reprise ciblée n'est pas négligeable en termes d'amélioration globale de la qualité du codage.

Sur l'ensemble de la population test de référence (IA + codage batch Iris + reprises manuelles, cf. Tableau 24), les niveaux de cohérence sont globalement très élevés : 95 % pour le score BS (moyenne des scores BLEU) et 86 % pour le score BS1 (moyenne des scores BLEU égaux à 1).

Le score BLEU est inférieur à 95 % pour cinq catégories :

- 2.1.01 – Tumeurs malignes des lèvres, de la bouche et du pharynx,
- 2.1.18 – Tumeurs malignes de la thyroïde
- 2.1.19 – Maladie de Hodgkin et lymphomes
- 17.1.1 – Accidents de transport
- 17.5 – Autres causes externes

La cohérence entre causes multiples reste inférieure à 85 % pour quatre catégories :

- 2.1.01 – Tumeurs malignes des lèvres, de la bouche et du pharynx
- 2.1.07 – Tumeurs malignes du larynx
- 2.1.19 – Maladie de Hodgkin et lymphomes
- 3 – Maladies du sang et des organes hématopoïétiques

Tableau 23. Cohérence des causes multiples prédites au niveau de la shortlist, sur la population de test de référence qui aurait été codée manuellement

	Part sur la population de test de référence qui aurait été codée manuellement dans une campagne							
		Transcription du processus morbide par IA		ajout de la reprise manuelle		ajout de la reprises cause multiple		
libele	nbObs	BS pondéré	BS1	BS pondéré	BS1	BS pondéré	BS1	Diff BS1 - BS1
01.1- Tuberculose	453	0,815	0,49	0,924	0,804	0,945	0,832	3%
01.2- SIDA (maladie VIH)	296	0,825	0,486	1	0,997	1	0,997	0%
01.3- Hepatites virales	340	0,773	0,388	0,907	0,762	0,921	0,779	2%
01.4- Autres maladies infectieuses et parasitaires	6428	0,841	0,537	0,928	0,799	0,944	0,822	2%
02.1.01-TM levre. bouche. pharinx	3014	0,807	0,496	0,864	0,633	0,89	0,672	4%
02.1.02-TM de l'oesophage	2642	0,881	0,628	0,912	0,717	0,923	0,734	2%
02.1.03-TM de l'estomac	2608	0,864	0,577	0,899	0,683	0,909	0,699	2%
02.1.04-TM Colon. rectum. anus	10785	0,87	0,578	0,903	0,674	0,915	0,694	2%
02.1.05-TM du foie et voies biliaires	5303	0,873	0,624	0,907	0,718	0,921	0,739	2%
02.1.06-TM du Pancreas	5996	0,887	0,635	0,919	0,728	0,93	0,746	2%
02.1.07-TM du Larynx	681	0,831	0,493	0,888	0,652	0,903	0,677	3%
02.1.08-TM du Poumon	17377	0,881	0,599	0,911	0,695	0,922	0,715	2%
02.1.09- Melanome malin de la peau	1298	0,869	0,605	0,897	0,687	0,916	0,713	3%
02.1.10-TM du sein	7608	0,886	0,597	0,913	0,689	0,92	0,7	1%
02.1.11-TM du col de l'uterus	563	0,853	0,561	0,902	0,691	0,915	0,714	2%
02.1.12-TM autre partie de l'uterus	1820	0,859	0,568	0,902	0,685	0,912	0,701	2%
02.1.13-TM de l'ovaire	1918	0,876	0,618	0,908	0,704	0,916	0,715	1%
02.1.14-TM de la prostate	5319	0,885	0,575	0,914	0,676	0,924	0,694	2%
02.1.15-TM du rein	2327	0,864	0,571	0,9	0,679	0,914	0,702	2%
02.1.16-TM de la vessie	3245	0,864	0,55	0,897	0,657	0,912	0,681	2%
02.1.17-TM cerveau et systeme nerveux central	2453	0,789	0,544	0,858	0,68	0,882	0,715	3%
02.1.18-TM de la thyroide	286	0,826	0,51	0,888	0,689	0,9	0,706	2%
02.1.19-Maladie de Hodgkin et lymphomes	3570	0,815	0,487	0,869	0,627	0,888	0,658	3%
02.1.20- Leucemie	3939	0,844	0,54	0,886	0,659	0,907	0,695	4%
02.1.21-Autres TM des tissus lymphoides et hematopoietiques	2155	0,86	0,546	0,898	0,661	0,91	0,682	2%
02.1.22-Autres Tumeurs malignes	16511	0,802	0,516	0,884	0,72	0,905	0,75	3%
02.2- Tumeurs non malignes	5768	0,828	0,507	0,909	0,736	0,926	0,764	3%
03- Maladies du sang et hematopoietiques	2326	0,831	0,505	0,899	0,695	0,918	0,725	3%
04.1- Diabete sucre	8021	0,879	0,52	0,918	0,662	0,926	0,676	1%
04.2- Autres maladies endocrinienness. nutritionnelles et metaboliques	6748	0,881	0,57	0,939	0,771	0,948	0,785	1%
05.1- Demences	8971	0,915	0,639	0,94	0,731	0,947	0,745	1%
05.2- Abus d'alcool	1727	0,87	0,576	0,95	0,834	0,96	0,851	2%
05.3- Pharmacodependance. toxicomanie	233	0,785	0,433	0,976	0,906	0,988	0,918	1%
05.4- Autres troubles mentaux et du comportement	2842	0,859	0,527	0,925	0,756	0,938	0,777	2%
06.1- Maladie de Parkinson	3226	0,914	0,624	0,938	0,72	0,945	0,733	1%
06.2 - Maladie d'Alzheimer	8616	0,902	0,631	0,929	0,725	0,938	0,74	2%
06.3- Autres maladies du systeme nerveux et des organes des sens	8274	0,851	0,543	0,917	0,743	0,934	0,768	3%
07.1.1-Infarctus aigu du myocarde	6894	0,887	0,618	0,92	0,723	0,931	0,74	2%
07.1.2-Autres Cardiopathies ischemiques	12151	0,88	0,539	0,917	0,676	0,926	0,69	1%
07.2-Autres maladies du coeur	25781	0,871	0,571	0,918	0,717	0,929	0,735	2%
07.3-Maladies cerebro vasculaires	20751	0,876	0,599	0,92	0,729	0,931	0,746	2%
07.4- Autres maladies de l'appareil circulatoire	16293	0,852	0,523	0,917	0,722	0,928	0,739	2%
08.1- Grippe	880	0,864	0,519	0,914	0,699	0,921	0,713	1%
08.2- Pneumonie	5024	0,891	0,65	0,932	0,776	0,945	0,795	2%
08.3.1- Asthme	468	0,86	0,513	0,921	0,718	0,929	0,731	1%
08.3.2-Autres maladies chroniques des voies respiratoires inferieures	6315	0,879	0,529	0,919	0,677	0,929	0,694	2%
08.4- Autres maladies de l'appareil respiratoire	7326	0,856	0,577	0,916	0,758	0,929	0,778	2%
09.1 - Ulcere gastro duodenal	670	0,844	0,5	0,909	0,736	0,924	0,76	2%
09.2 - Cirrhoses. fibroses et hepatites chroniques	4497	0,882	0,582	0,919	0,714	0,928	0,728	1%
09.3- Autres maladies de l'appareil digestif	12472	0,834	0,499	0,907	0,705	0,922	0,729	2%
10- Maladies de la peau et du tissu sous-cutane	1382	0,823	0,509	0,934	0,8	0,946	0,816	2%
11.1- Arthrite rumathoide et osteroarthrite	495	0,849	0,499	0,935	0,778	0,944	0,792	1%
11.2- Autres maladies du systeme osteoarticulaire et des muscles	3370	0,764	0,369	0,925	0,784	0,938	0,804	2%
12.1-Maladies du rein et de l'uretre	4889	0,838	0,529	0,91	0,741	0,924	0,763	2%
12.2- Autres maladies genito-uniraires	2598	0,782	0,457	0,888	0,709	0,909	0,741	3%
13- Complications de grossesse et d'accouchement	59	0,492	0,153	1	1	1	1	0%
14- Certains affections dont l'origine est perinatale	1854	0,684	0,328	1	0,999	1	0,999	0%
15- Malformations congenitales et anomalies chromosomiques	1492	0,702	0,326	0,938	0,838	0,954	0,863	3%
16.1- Syndrome de mort subite nourisson	188	0,971	0,91	0,997	0,984	0,997	0,984	0%
16.2- Causes inconnues ou non precisees	5513	0,887	0,736	0,946	0,877	0,956	0,893	2%
16.3- Autres symptomes et etat morbides non definis	7349	0,917	0,786	0,944	0,851	0,957	0,872	2%
17.1.1- Accidents de transport	2325	0,753	0,47	0,834	0,637	0,907	0,751	11%
17.1.2 - Chutes accidentelles	10474	0,883	0,569	0,954	0,83	0,96	0,841	1%
17.1.3 - Noyade et submersion accidentelle	416	0,758	0,519	0,875	0,764	0,892	0,793	3%
17.1.4 - Intoxications accidentelles	1845	0,783	0,463	0,947	0,872	0,96	0,889	2%
17.1.5 - Autres accidents	15489	0,837	0,58	0,935	0,83	0,948	0,85	2%
17.2- Suicide	5059	0,751	0,509	0,826	0,653	0,896	0,756	10%
17.3- Homicides	400	0,598	0,308	0,924	0,858	0,974	0,927	7%
17.4- Evenement dont l'intention n'est pas determinee	1671	0,698	0,47	0,926	0,879	0,96	0,923	4%
17.5- Autre cause externe	1672	0,721	0,388	0,909	0,795	0,925	0,822	3%
18- Covid	14895	0,877	0,548	0,907	0,654	0,914	0,665	1%
Total	**368644**	**0,858**	**0,564**	**0,915**	**0,728**	**0,928**	**0,749**	

Tableau 24. Cohérence des causes multiples prédites au niveau de la shortlist sur la campagne complétée à 3 modes de codages

	Population test de référence complète, campagne qui combine les trois modes de codage						
		Transcription du processus morbide par IA et par batch Iris		ajout de la reprise manuelle		ajout de la reprises cause multiple	
libele	nbObs	BS pondéré	BS1	BS pondéré	BS1	BS	BS1
01.1- Tuberculose	507	0,835	0,544	0,932	0,824	0,951	0,85
01.2- SIDA (maladie VIH)	379	0,863	0,599	1	0,997	1	0,997
01.3- Hepatites virales	597	0,871	0,652	0,947	0,864	0,955	0,874
01.4- Autres maladies infectieuses et parasitaires	15471	0,934	0,808	0,97	0,916	0,977	0,926
02.1.01-TM levre. bouche. pharinx	5707	0,898	0,734	0,928	0,806	0,942	0,827
02.1.02-TM de l'oesophage	5570	0,944	0,823	0,958	0,866	0,963	0,874
02.1.03-TM de l'estomac	6764	0,948	0,837	0,961	0,878	0,965	0,884
02.1.04-TM Colon. rectum. anus	26880	0,948	0,831	0,961	0,869	0,966	0,877
02.1.05-TM du foie et voies biliaires	13408	0,95	0,851	0,963	0,888	0,969	0,897
02.1.06-TM du Pancreas	18506	0,963	0,882	0,974	0,912	0,977	0,918
02.1.07-TM du Larynx	1422	0,919	0,757	0,947	0,833	0,954	0,845
02.1.08-TM du Poumon	47508	0,956	0,853	0,967	0,889	0,972	0,896
02.1.09- Melanome malin de la peau	2636	0,935	0,805	0,949	0,846	0,959	0,859
02.1.10-TM du sein	19581	0,956	0,843	0,966	0,879	0,969	0,884
02.1.11-TM du col de l'uterus	1223	0,932	0,798	0,955	0,858	0,961	0,868
02.1.12-TM autre partie de l'uterus	4265	0,94	0,815	0,958	0,866	0,962	0,872
02.1.13-TM de l'ovaire	5164	0,954	0,858	0,966	0,89	0,969	0,894
02.1.14-TM de la prostate	13857	0,956	0,837	0,967	0,876	0,971	0,883
02.1.15-TM du rein	5303	0,94	0,812	0,956	0,859	0,962	0,869
02.1.16-TM de la vessie	7986	0,945	0,817	0,958	0,861	0,964	0,87
02.1.17-TM cerveau et systeme nerveux central	6196	0,916	0,82	0,944	0,873	0,953	0,887
02.1.18-TM de la thyroide	568	0,912	0,754	0,944	0,843	0,949	0,852
02.1.19-Maladie de Hodgkin et lymphomes	7440	0,911	0,754	0,937	0,821	0,946	0,836
02.1.20- Leucemie	9203	0,933	0,803	0,951	0,854	0,96	0,869
02.1.21-Autres TM des tissus lymphoides et hematopoietiques	4977	0,939	0,803	0,956	0,853	0,961	0,862
02.1.22-Autres Tumeurs malignes	34170	0,904	0,766	0,944	0,865	0,954	0,879
02.2- Tumeurs non malignes	11784	0,916	0,759	0,955	0,871	0,964	0,885
03- Maladies du sang et hematopoietiques	4166	0,906	0,724	0,944	0,83	0,954	0,846
04.1- Diabete sucre	18601	0,948	0,793	0,965	0,854	0,968	0,86
04.2- Autres maladies endocrinienness. nutritionnelles et metaboliques	16630	0,952	0,825	0,975	0,907	0,979	0,913
05.1- Demences	29165	0,974	0,889	0,981	0,917	0,984	0,921
05.2- Abus d'alcool	3818	0,941	0,808	0,978	0,925	0,982	0,933
05.3- Pharmacodependance. toxicomanie	364	0,863	0,637	0,985	0,94	0,992	0,948
05.4- Autres troubles mentaux et du comportement	5822	0,931	0,769	0,963	0,881	0,97	0,891
06.1- Maladie de Parkinson	10473	0,973	0,884	0,981	0,914	0,983	0,918
06.2 - Maladie d'Alzheimer	29207	0,971	0,891	0,979	0,919	0,982	0,923
06.3- Autres maladies du systeme nerveux et des organes des sens	18562	0,934	0,796	0,963	0,886	0,971	0,897
07.1.1-Infarctus aigu du myocarde	20834	0,963	0,873	0,974	0,908	0,977	0,914
07.1.2-Autres Cardiopathies ischemiques	28348	0,948	0,802	0,965	0,861	0,968	0,867
07.2-Autres maladies du coeur	78522	0,958	0,859	0,973	0,907	0,977	0,913
07.3-Maladies cerebro vasculaires	48234	0,947	0,828	0,966	0,884	0,97	0,891
07.4- Autres maladies de l'appareil circulatoire	38266	0,937	0,797	0,965	0,882	0,969	0,889
08.1- Grippe	2092	0,943	0,798	0,964	0,873	0,967	0,879
08.2- Pneumonie	19661	0,972	0,911	0,983	0,943	0,986	0,948
08.3.1- Asthme	1238	0,947	0,816	0,97	0,893	0,973	0,898
08.3.2-Autres maladies chroniques des voies respiratoires inferieures	15469	0,951	0,808	0,967	0,868	0,971	0,875
08.4- Autres maladies de l'appareil respiratoire	25051	0,958	0,876	0,975	0,929	0,979	0,935
09.1 - Ulcere gastro duodenal	1292	0,919	0,741	0,953	0,863	0,961	0,875
09.2 - Cirrhoses. fibroses et hepatites chroniques	10485	0,949	0,821	0,965	0,878	0,969	0,884
09.3- Autres maladies de l'appareil digestif	26210	0,921	0,762	0,956	0,86	0,963	0,871
10- Maladies de la peau et du tissu sous-cutane	2482	0,901	0,726	0,963	0,888	0,97	0,898
11.1- Arthrite rumathoide et osteroarthrite	865	0,914	0,713	0,963	0,873	0,968	0,881
11.2- Autres maladies du systeme osteoarticulaire et des muscles	5177	0,846	0,589	0,951	0,859	0,96	0,872
12.1-Maladies du rein et de l'uretre	12548	0,937	0,816	0,965	0,899	0,97	0,908
12.2- Autres maladies genito-uniraires	4851	0,883	0,709	0,94	0,844	0,951	0,861
13- Complications de grossesse et d'accouchement	62	0,516	0,194	1	1	1	1
14- Certains affections dont l'origine est perinatale	2192	0,732	0,432	1	0,999	1	0,999
15- Malformations congenitales et anomalies chromosomiques	2373	0,813	0,576	0,961	0,898	0,971	0,914
16.1- Syndrome de mort subite nourisson	193	0,972	0,912	0,997	0,984	0,997	0,984
16.2- Causes inconnues ou non precisees	25097	0,975	0,942	0,988	0,973	0,99	0,976
16.3- Autres symptomes et etat morbides non definis	49164	0,988	0,968	0,992	0,978	0,994	0,981
17.1.1- Accidents de transport	3973	0,855	0,69	0,903	0,788	0,946	0,854
17.1.2 - Chutes accidentelles	13336	0,908	0,662	0,964	0,866	0,969	0,875
17.1.3 - Noyade et submersion accidentelle	1232	0,918	0,838	0,958	0,92	0,963	0,93
17.1.4 - Intoxications accidentelles	2437	0,836	0,594	0,96	0,903	0,97	0,916
17.1.5 - Autres accidents	21429	0,882	0,696	0,953	0,877	0,962	0,892
17.2- Suicide	12480	0,899	0,801	0,93	0,859	0,958	0,901
17.3- Homicides	555	0,71	0,501	0,946	0,897	0,981	0,948
17.4- Evenement dont l'intention n'est pas determinee	2108	0,761	0,58	0,941	0,904	0,968	0,939
17.5- Autre cause externe	1994	0,766	0,487	0,923	0,828	0,937	0,851
18- Covid	41739	0,956	0,839	0,967	0,876	0,969	0,88
Total	**935869**	**0,944**	**0,828**	**0,967**	**0,893**	**0,972**	**0,901**

8.11 Constitution de la population test de référence

Pour l'évaluation de la campagne de l'année de 2023, on repart de la base de données constituée pour l'évaluation de la campagne de 2022[11], nommée « population test de référence » et on y ajoute des données issues de la campagne 2023. Toutes les données codées manuellement sont nécessairement issues du champ « test » de l'IA, c'est-à-dire que ce sont des certificats qui n'ont pas servi à l'entraînement des modèles, et qui permettent donc l'évaluation des modèles et ici l'évaluation de la campagne.

8.11.1 Quelles données sont disponibles pour la campagne 2023 ?

En décembre 2024, avant la synchronisation des données CépiDc de 2023 avec celles de l'Insee, 639 007[12] certificats avaient été reçus. Parmi ceux-là, 405 690 ont été codés *via* le système expert IRIS/Muse (63,5 %). Il restait donc 233 317 certificats à coder, soit manuellement, soit *via* prédiction IA.

Dans le cadre de la campagne 2023, 15 échantillons ont été prélevés au cours de la campagne : ce sont ces échantillons qui vont être codés manuellement pour la campagne 2023 (en plus des vérifications et éventuellement reprises de certains certificats). On se restreint donc aux 15 échantillons prélevés en 2023.

Tableau 25 – Organisation de la campagne 2023

Echantillon	Effectifs total	dont manuel	Effectifs test*	dont manuel
Données non-échantillonnées traitées par codage automatique (batch IRIS/Muse)	343 807			
Données non-échantillonnées traitées par IA	135 419	-	-	-
ECH001 - Aléatoire sur l'ensemble de 2023 ECH002 - Champ EDP	68 200 28 555	24 678 10 194	13 573 28 552	13 573 10 191
ECH003 - Aléatoire sur les non-automatiques pas encore tirés	7 000	7 000	5 248	5 248
ECH005 - IA ciblé 94%	9 785	9 785	978	978
ECH006 - IA ciblé 95%	4 692	4 692	469	469
ECH008 - Décès sensibles	2 890	2 890	275	275
ECH009 - IA ciblé 96%	7 950	7 950	788	788
ECH010 - IA ciblé 96,5%	6 067	6 067	589	589
ECH011 - IA ciblé 97%	6 845	6 845	663	663
ECH012 - Aléatoire sur les non-automatiques pas encore tirés dans ECH001/002	5 000	5 000	4 914	4 914

[11] Pour un récapitulatif de la création de la population test de référence pour la production des causes de décès 2022, voir Annexe 11 du rapport 2022.

[12] La base de test a été réalisée avant synchronisation avec l'Insee donc notamment avant dédoublonnage et avant intégration des certificats arrivés tardivement.

*ECH018 - Aléatoire sur les non-automatiques pas encore tirés dans ECH001/002/012 ** *	2 500	2 500	-	-
ECHrepriseCA - P1	2 574	2 574	749	749
ECHrepriseCA - P2	2 574	2 574	747	747
ECHrepriseCA - P3	2 574	2 574	764	772
*ECHrepriseCA - P4** *	2 575	2 575	-	-
Total	639 007	97 898	58 309	39 956

* Ne sont compris que les certificats identifiés comme « » test » » **et** toujours présents dans les données après synchronisation, afin que la population test de référence ne soit pas constituée en partie de certificats non-conservés par le CépiDc.

** Les certificats présents dans les échantillons « » *ECH018* » » et « » *ECHrepriseCA - P4* » », initialement tirés pour être codés manuellement, ont finalement été codés par IA.

8.11.2 Comment ont-été tirés les échantillons de la campagne 2023 ?

Un premier échantillon (ECH001) représentant 10,7 % de décès de 2023 est prélevé aléatoirement dans l'ensemble des certificats de décès.

Un deuxième échantillon (ECH002) est constitué des certificats de décès des personnes appartenant à l'Echantillon Démographique Permanent[13]. En tant que champ représentatif de la population vivante, il est aussi représentatif de la population des décédés.

Quatre autres échantillons (ECH005, ECH006, ECH010 et ECH011) rassemblent des certificats tirés spécifiquement pour augmenter la précision du codage par intelligence artificielle sur certaines catégories de maladies précises (celles pour lesquelles l'IA a des scores de précision inférieurs) : en réorientant des certificats identifiés au préalable comme particulièrement difficiles vers le codage manuel, la précision de l'IA sur ces maladies augmente mécaniquement.

Un échantillon (ECH008) est constitué de l'ensemble des certificats présentant des décès pour des causes dites « sensibles » comme les morts maternelles, les morts mentionnant VIH/SIDA et presque tous les décès avant 15 ans.

Trois échantillons (ECH003, ECH012 et ECH018) sont tirés aléatoirement sur les certificats restant après avoir éliminé les certificats codés par système expert et ceux déjà isolés dans un échantillon aléatoire. ECH018 prévu pour être codé manuellement, a finalement été codé par IA.

Enfin, les 4 derniers échantillons (ECHrepriseCA - P1, P2, P3 et P4) ont été tirés pour la première fois en 2023. Ils visent à mesurer et améliorer la qualité de la retranscription de l'entièreté du processus morbide par l'imputation IA et non simplement la cause initiale retenue. Un seul de ces échantillons prévus pour être codé manuellement, a finalement été codé par IA.

[13] Le champ EDP est représentatif de la population en ce que naître un jour « EDP » correspond à un tirage aléatoire en population générale, on peut donc reproduire ce tirage pour obtenir le même effet. Pour rappel, appartiennent au champ EDP toute personne née les 2,3,4,5 janvier ou les 1,2,3,4 avril, juillet, octobre, voir : L'échantillon démographique permanent : en 50 ans, l'EDP a bien grandi ! Isabelle Robert-Bobée, Natacha Gualbert, Courrier des statistiques N°6 – 2021.

8.11.3 Compléter la population test de référence

On part donc de la population test de référence utilisée pour évaluer la campagne 2022 et on souhaite la compléter avec des données issues de 2023. Pour conserver la distribution des causes de décès en 2023, il est essentiel de n'utiliser que des échantillons aléatoires ou représentatifs de la population générale décédée : aussi, seuls les échantillons ECH001, ECH002, ECH003 et ECH012 seront donc utilisés pour compléter la base en certificats codés manuellement. De même, seule la partie « test » de ces échantillons est mobilisable pour évaluer la campagne. On viendra ensuite tirer des certificats codés automatiquement pour obtenir la même proportion de codage automatique dans l'ajout à la population test de référence que dans la campagne 2023 (63,5 %).

Données EDP (ECH002) :

L'ECH002, constitué des certificats de décès des personnes appartenant au champ EDP, est entièrement mis dans la partie « test ». De même, l'échantillon possède déjà la répartition attendue entre codage automatique par batch et codage manuel. Il est donc intégré dans sa totalité à la population test de référence.

Données tirées aléatoirement (ECH001, ECH003, ECH012) :

Les trois échantillons étant tirés aléatoirement, on agrège ensemble les certificats manuels « tests » : 13573 + 5248 + 4914 = 23 735. On a donc 23 735 certificats codés manuellement et on tire dans le codage automatique la bonne proportion de certificats pour obtenir la même répartition (36,5 %/63,5 %). On fait attention à exclure les certificats appartenant au champ EDP, déjà pris en compte. Le calcul suivant donne le nombre de certificats à tirer :

$$23\,735 \times \frac{(certificats\ IRIS/MUSE\ - EDP)}{(certificats\ pas\ IRIS/MUSE - EDP)} = nombre\ de\ certificats\ IRIS/MUSE\ \text{à}\ tirer$$

Donc $$23\,735 \times \frac{405\,690}{223\,126} = 40\,908\ certificats$$

On prélève donc 40 908 certificats dans les certificats codés par le batch automatique hors du champ de l'échantillon EDP.

On obtient alors la composition de la population test de référence suivante :

Tableau 26 – Répartition des types de codage dans la base de test de référence finale.

Année	Type de codage	Effectifs	Répartition (en %)
2016	Automatique	128 663	58
	Manuel	93 144	42
2017	Automatique	132 944	58,6
	Manuel	93 912	41,4
2020	Automatique	164 323	57,5
	Manuel	121 461	42,5
2021	Automatique	39 109	62

	Manuel	23 949	38
2022	Automatique	42 488	62,5
	Manuel	25 491	37,5
2023	Automatique	59 269	63,6
	Manuel	33 926	36,4
Total	Total	958 679	
	Dont Automatique	567 225	59,2
	Dont Manuel	391 454	40,8

La base de population test de référence finale est donc composée de 958 679 certificats, dont 391 883 ont été codés manuellement durant les productions de 2016, 2017, 2020, 2021, 2022 et 2023 et n'ont jamais été utilisés pour entraîner les modèles d'IA. Les certificats codés automatiquement peuvent avoir été utilisés pour l'entraînement de l'IA, car il n'y aura pas d'évaluation du codage IA *vs* codage automatique par IRIS/Muse[14] : ils servent simplement à établir le niveau global de cohérence, donc batch automatique y compris.

On observe que la base de population test de référence a, en moyenne, moins de codage automatique que la base de production de 2023 (voir Tableau 26) : cela résulte d'un arbitrage entre conserver la part adéquate de codage automatique et conserver la répartition du pourcentage de codage automatique pour chaque cause de décès chaque année. En effet, le dictionnaire et les règles de traitement qui permettent le fonctionnement du système expert peuvent changer de campagnes en campagnes (particulièrement entre 2020 et 2021 avec un important travail sur la standardisation des données réalisé lors du « rattrapage » des années 2018-2019), ce qui a pour conséquence directe de modifier la répartition des causes de décès codées automatiquement ou non. Chercher à obtenir la même part de codage automatique dans la population test de référence que dans la campagne 2023 aurait pour effet de déformer la distribution des causes de décès et introduire des biais très difficilement estimables dans l'évaluation suivante. On préfère donc un taux de codage automatique inférieur dans la population test de référence (par rapport à 2023), qui implique uniquement une légère sous-estimation de la cohérence du codage dans les analyses présentées ci-dessous[15].

De même, on observe que la population test de référence est composée à 51,8 % de données de 2016-2017, contre 15,1 % pour les données de 2021-2022 : la distribution globale des causes de décès de la population de référence est donc biaisée vers la distribution des plus anciennes années. Cela implique une sous-estimation de la cohérence du codage pour les statistiques toutes causes confondues[16]. Une étude sur le biais provoqué par l'évolution de la distribution des causes de décès entre 2016-2017 et 2021-2023 reste à réaliser.

[14] Le codage automatique est entièrement comparé à lui-même, car la prédiction par IA n'est réalisée qu'après que le certificat n'ait pu être automatiquement codé par IRIS/Muse : il n'y a donc aucune situation où la prédiction IA remplace le codage automatique.

[15] Puisque le codage automatique se compare à lui-même, sa cohérence sera parfaite pour chaque certificat codé automatiquement. Mécaniquement, moins de certificats codés automatiquement implique une baisse de la cohérence générale, tous les autres types de codage pouvant être incohérents.

[16] Le tableau XX nous montre que la campagne est plus cohérente sur les données récentes.

8.12 Liste des tableaux

Tableau 1. Exhaustivité de la collecte des volets médicaux (VM) pour les décès survenus entre 2021 et 2023 5
Tableau 2. Exhaustivité de la collecte des volets médicaux (VM) pour les décès survenus entre 2021 et 2023 7
Tableau 3. Effectifs et répartition des certificats reçus en 2023, 2022 et 2021, par mode de codage..... 11
Tableau 4. Nombre de certificats repris manuellement après un ciblage IA sur la cause initiale lors de la campagne de codage 2023, selon le niveau de précision à atteindre pour chaque catégorie de la shortlist européenne 12
Tableau 5. Répartition des codages manuels lors de la campagne 2023, selon l'origine du certificat 15
Tableau 6. Décès identifiés comme sensibles en 2023, selon la source de sensibilité..... 16
Tableau 7. Description et nombre de certificats vérifiés pour les décès spécifiques et part de cause initiale modifiées..... 23
Tableau 8. Vérifications liées au système expert, par motif et parts de cause initiale et causes associées modifiées..... 24
Tableau 9. Vérifications liées aux nouvelles règles, par motif et parts de cause initiale et causes associées modifiées..... 25
Tableau 10. Vérifications liées à la démarche de choix de code, par motif et parts de cause initiale et causes associées modifiées..... 26
Tableau 11. Description de la distribution des allocations de codage dans la base 2023 et sa simulation dans la population test de référence..... 29
Tableau 12. Cohérence (accuracy) entre causes initiales prédites par deep learning (K5), combinaison de deep learning et Iris Muse (K5 iris), proposition du surmodèle, avec ou sans prise en compte de la reprise manuelle 2023 pour les données qui auraient été codées manuellement dans une campagne classique de codage..... 30
Tableau 13. Cohérence entre causes initiales prédites par deep learning (K5), combinaison de deep learning et Iris Muse (K5Iris), proposition du surmodèle, avec ou sans prise en compte de la reprise manuelle 2023 pour l'ensemble de la population (et 2022 pour référence) 31
Tableau 14. Performances du sur-modèle et du sur-modèle combiné avec la reprise manuelle évalués sur les observations de la population test qui auraient été codées manuellement. 32
Tableau 15. Performances du système expert combiné au sur-modèle et au sur-modèle combiné et à la reprise manuelle évaluées sur l'ensemble de la population test de référence, par item de la shortlist Eurostat. 33
Tableau 16. Performances en termes de cohérence de chapitre de la CIM-10 de l'ensemble de la campagne 2023 et effectifs prédits sur l'ensemble de la population test de référence. 35
Tableau 17. Evaluation sur la population test des gains en cohérence (accuracy) des différentes étapes de reprises manuelles telles qu'elles ont été menées pour la campagne 2023 36
Tableau 18 Effectifs à reprendre manuellement pour atteindre XX % de précision dans la catégorie 45
Tableau 19. Effectifs de la base d'apprentissage 65
Tableau 20. Proportion des types de proposition en base 66
Tableau 21. Score BLEU pondéré des causes multiples prédites par le modèle de deep learning (K5), ainsi que des résultats obtenus après combinaison des prédictions du modèle avec les reprises manuelles, d'abord sur la qualité de la cause initiale, puis sur celle des causes multiples, comparées aux données qui auraient été codées manuellement dans une campagne traditionnelle. 84
Tableau 22. Score BLEU pondéré des causes multiples obtenu en combinant les deux modes de codage deep learning (K5) et batch, ainsi que sur l'ensemble de la campagne de codage intégrant les trois modes. 85
Tableau 23. Cohérence des causes multiples prédites au niveau de la shortlist, sur la population de test de référence qui aurait été codée manuellement 87
Tableau 24. Cohérence des causes multiples prédites au niveau de la shortlist sur la campagne complétée à 3 modes de codages..... 88
Tableau 25 – Organisation de la campagne 2023..... 89
Tableau 26 – Répartition des types de codage dans la base de test de référence finale. 91

8.13 Liste des figures

Figure 1. Répartition de l'ensemble des décès et des décès manquants par groupe d'âge en 2023 8
Figure 2. Stratégie de combinaison des trois modes de codage 11
Figure 3. Codage par système expert Iris/Muse 14
Figure 4. Exemple de séquences d'entrée et de sortie du modèle de neurones transformer pour coder la séquence de causes (modèle dit « K5 ») 17
Figure 5. Description des propositions possibles de causes initiales impliquant du codage par IA 20
Figure 6. Distribution du score BLEU réel dans les bases de Train et de Test 48
Figure 7. Architecture du modèle Transformers k5 implémenté pour le codage des causes de décès par AI 60
Figure 8. Répartition selon Origine - test 2023 (381 292 obs) 70