

Les données des certificats de décès en France : processus de production et principaux types d'analyse

Death certificate data in France: production process and main types of analyzes

Grégoire Rey

CépiDc, Inserm, Le Kremlin-Bicêtre, France

Grégoire Rey

gregoire.rey@inserm.fr

CépiDc, Inserm,

80, rue du Général Leclerc, 94276 Le Kremlin-Bicêtre CEDEX

Version Archives ouvertes de l'article publié dans La Revue de Médecine Interne, Volume 37, Issue 10, October 2016, Pages 685-693

Résumé

Introduction

Les données de mortalité, par le peu d'ambiguïté de leur définition et de leur compréhension par l'ensemble des acteurs, et l'exhaustivité de leur enregistrement, constituent une pierre angulaire de la statistique en santé publique en France et dans la plupart des pays industrialisés. Cet article décrit le processus de production des données, et les principaux types d'analyses possibles.

Processus de production des données

La production des données se décompose en différentes étapes : la certification par le médecin sur support papier ou électronique (à l'aide d'une application web), la transmission des données à l'Inserm, la saisie et le codage de l'information.

Le codage de l'information vise à suivre les recommandations de l'OMS formulées dans la Classification Internationale des Maladies (CIM, 10^{ème} révision utilisée depuis 2000). Il est effectué à l'aide d'un logiciel de codage automatique, nommé Iris, développé au sein d'un consortium international. Le codage consiste en premier lieu à attribuer un code de la CIM à toutes les entités nosologiques rencontrées sur le certificat, puis à choisir la cause initiale de décès. Cette dernière est la principale information utilisée pour les exploitations statistiques.

Principaux types d'analyses

Trois principaux types d'analyses se dégagent dans la littérature : l'exploitation des données présentes sur le certificat de décès uniquement, les analyses écologiques (études d'associations entre des variables mesurées à l'échelle de groupes) et les analyses à partir de données chaînées individuellement à d'autres bases de données. Ces différentes analyses permettent de traiter de nombreuses problématiques en santé publique.

Perspectives

Plusieurs évolutions dans le processus de production sont en cours de mise en œuvre : le déploiement de la certification électronique, l'automatisation accrue du traitement de l'information contenue sur les certificats et le chaînage pérenne et complet de l'information avec les données de l'assurance maladie et d'hospitalisation.

Elles pourraient prochainement encore profondément élargir le champ des exploitations possibles des données de causes de décès.

Mots clés :

Causes de décès, Statistique nationale, Codage médical, Santé publique

Abstract

Introduction

Mortality data, by the unambiguity of their definition and understanding by all stakeholders, and completeness of registration, are a cornerstone of public health statistics in France and in most industrialized countries. This article describes the data production process, and the main types of possible analyzes.

Process of data production

Data production is composed of different stages: death certification by a medical doctor on paper or electronic (using a web application) format, data transmission to Inserm, capture and coding of information.

The encoding of the information follows the WHO recommendations of the International Classification of Diseases (ICD, 10th revision used since 2000). It is carried out using an automatic coding software, called Iris, developed in an international consortium. The coding aims, first, at assigning an ICD code to all nosologic entities encountered on the certificate, and then at selecting the underlying cause of death. The latter is the main information used for statistical analyzes.

Main types of analyzes

Three main types of analysis emerge in the literature: the exploitation of data on the death certificate only, ecological analyzes (studies of associations between variables measured across groups) and analysis from data individually linked to other databases. Many public health issues can be addressed with these various analyzes.

Perspectives

Several developments in the production process are being implemented: the deployment of electronic certification, increased automation of the death certificate information processing and durable and complete record linkage with health insurance and hospitalization data.

They could soon be deeply expanding the scope of possible uses of causes of death data.

Keywords :

Causes of death, National Statistics, Medical coding, Public Health

Introduction

Les données de mortalité figurent parmi les plus anciennes données d'observation de la santé humaine existantes. Ceci est dû à l'universalité de la mort, la facilité de son observation, mais aussi à ses conséquences administratives importantes qui doivent être régies par la communauté. De ce fait, elle est l'un des événements de la vie dont l'enregistrement est le plus systématique et le plus exhaustif.

Historiquement, lors de son enregistrement dans les registres paroissiaux, puis à l'état civil, la mortalité a été d'emblée associée à la notion de cause, ou de circonstances de décès [1]. Mais ce n'est qu'à la fin du 19^{ème} siècle que la déclaration des causes de décès en France a donné lieu à une formalisation et une classification permettant une exploitation statistique rigoureuse [2]. La statistique de mortalité par cause a d'abord été produite dans les mêmes conditions que les données d'état civil par l'INSEE, et essentiellement utilisées par des démographes. Cependant, depuis 1968, le Code Général des Collectivités Territoriales (article L2223-42) a assigné à l'Inserm cette mission de production, assortie d'une mission de recherche pour la santé publique à partir de ces données. Le caractère médical et confidentiel de ces données a ainsi amené le législateur à considérer que leur gestion et leur exploitation seraient plus appropriées au sein d'un organisme fortement lié à l'univers médical. Aujourd'hui, les données sont produites au sein d'une unité de service de l'Inserm, le Centre d'épidémiologie sur les causes médicales de décès (CépiDc) en collaboration avec le Département Système d'Information (DSI) de l'Inserm. Elles sont exploitées le plus souvent par des épidémiologistes dans un contexte nécessitant des compétences à la fois démographiques et médicales.

A partir du moment où les analyses ont été facilitées par les progrès de l'informatisation, ces données ont été de plus en plus exploitées pour des objectifs de santé publique. Cette orientation s'est traduite par leur intégration dans de nombreux indicateurs de la loi de santé publique de 2004 [3].

Ces indicateurs, obtenus à partir de données observationnelles en routine, sont rarement considérés comme permettant de mettre en évidence avec un fort niveau de preuve l'effet d'un facteur causal sur l'état de santé d'une population en général, et sur la survenue de maladies spécifiques. Ils utilisent les informations recueillies sur le sujet décédé lors de la déclaration du décès à l'état civil (sexe, âge, lieu de domicile et de décès, profession...). Ces données sont obtenues en dehors d'un contexte déclaratif garantissant la parfaite pertinence de l'information recueillie. En effet les causes de décès sont le plus souvent déclarées par le médecin constatant le décès, qui ne connaît pas toujours le patient, ni l'historique des raisons de sa mort. Enfin, la mortalité demeure un événement

extrême ne permettant pas d'appréhender l'ensemble de l'état de santé d'une population, et en particulier pas le niveau de prévalence d'affections non ou peu létales, même si elles sont très invalidantes.

Malgré ces limites, les données de causes de décès par leur caractère exhaustif, historiquement et internationalement comparable, représentent pour de nombreuses problématiques une information de référence permettant d'apporter un éclairage sur les principaux enjeux de santé publique et d'évaluer l'effet de nombreux facteurs sur la santé. Le CépiDc enregistre chaque année de nombreuses demandes de données émanant d'institutions nationales (Inserm, INED, InVS, ORS, ARS, IRSN, INRS, DREES, Universités...) et internationales (OMS, Eurostat, OCDE). En particulier, l'analyse descriptive à partir de données écologiques a été à l'origine de la formalisation d'hypothèses qui ont pu être testées dans un deuxième temps à l'aide de protocoles plus rigoureux. De plus, la multiplication actuelle des possibilités d'appariement de ces données avec d'autres bases de données médicales, administratives, démographiques ou épidémiologiques ouvre de nombreuses perspectives d'enrichissement et d'analyse.

L'objet de cet article est de présenter les modalités de production des données issues des certificats de décès, les modalités d'accès aux données et les principaux types d'analyses qui peuvent être mises en œuvre à partir de ces données.

Production des données

L'article L. 2223-42 du Code Général des Collectivités Territoriales stipule que "L'autorisation de fermeture du cercueil ne peut être délivrée qu'au vu d'un certificat, établi par un médecin, attestant le décès. Ce certificat, rédigé sur un modèle établi par le ministère chargé de la santé, précise la ou les causes de décès, aux fins de transmission à l'Institut national de la santé et de la recherche médicale."

Le circuit du certificat de décès est le même pour toutes les personnes décédées en France, qu'elles soient françaises ou non, résidant en France ou non.

Cependant, il existe depuis 2007 deux modalités pour établir un certificat de décès (Figure 1) :

- la certification électronique des décès,
- la certification papier.

Le système de certification électronique fonctionne via une application web. Ce système monte lentement en charge. Il représentait 0,9% des décès en 2008, 5% en 2012 et en 2015 il atteint

environ 10% des décès. Nous décrivons ci-dessous en premier lieu le circuit de l'information associé à une certification papier. Puis nous décrivons les spécificités du circuit de l'information associées à une certification électronique.

Figure 1 ici

[Le médecin qui constate le décès remplit le certificat de décès](#)

Le certificat de décès (CDC) comprend deux parties (figure 2) :

- La partie supérieure destinée au bureau de l'état civil de la mairie de la commune de décès, qui est nominative et qui contient des informations administratives
- La partie inférieure destinée à l'analyse des causes médicales de décès, qui est anonyme et cachetée par le médecin certificateur et contient des informations administratives et médicales relatives au défunt, le cachet et la signature du médecin.

L'espace dédié aux causes médicales de décès est conforme au modèle international préconisé par l'OMS dans la Classification Internationale des Maladies (CIM) et comprend lui-même deux parties :

- La partie 1 qui comporte 4 lignes qui permettent au médecin de décrire l'enchaînement causal des maladies ayant directement conduit à la mort, de la cause immédiate rapportée sur la première ligne à la cause initiale mentionnée sur la dernière ligne remplie.
- La partie 2 qui permet de notifier les autres états morbides ayant pu contribuer au décès.

Pour les décès survenant avant 28 jours de vie, un certificat néonatal de décès est établi en lieu et place du certificat de décès. Il contient des informations supplémentaires relatives aux parents du défunt (profession, nationalité, statut matrimonial et de concubinage, nombre de grossesses et d'accouchements de la mère) et à l'accouchement (présentation, début du travail, mode (voie basse, extraction opératoire par voie basse, césarienne)).

Figure 2 ici

[Le certificat est transmis du médecin certificateur à la mairie du lieu de décès \(dans un délai de 48h\)](#)

Un agent d'état civil :

- enregistre la déclaration de décès et complète le numéro d'acte et d'ordre. La mairie conserve la partie supérieure du certificat de décès,

- établit l'avis 7 bis (nominatif) et le bulletin 7 (copie anonyme de l'avis 7 bis). Les informations d'état civil sont enregistrées à l'aide des informations présentes sur la partie supérieure du certificat de décès et d'informations complémentaires (la commune ou le pays de naissance, la profession, le statut matrimonial) obtenues auprès de la famille, de l'administration hospitalière, de l'organisme funéraire, de l'acte de naissance ou du livret de famille. L'avis 7 bis est transmis à l'INSEE, où il vient mettre à jour le Répertoire National d'Identification des Personnes Physiques (RNIPP), en y apportant, en plus du Numéro d'Identification au Répertoire (NIR), des nom, prénom, sexe, date et commune de naissance déjà présents, la date et la commune de décès.

Le bulletin 7 et la partie inférieure du CDC sont transmis à la délégation territoriale de l'ARS, puis au CépiDc

La partie inférieure du CDC associée au bulletin 7 est envoyée par la mairie du lieu de décès à la délégation territoriale du département de décès dépendant de l'ARS. La partie médicale y est décachetée par le médecin inspecteur de santé publique qui prend connaissance des causes de décès dans l'éventualité de mesures de sauvegarde, mais aucune information n'est modifiée. Ces documents sont ensuite transmis au CépiDc, où la partie inférieure du CDC et le bulletin 7 sont scannés et saisis en conservant l'association entre les deux documents.

Spécificités de la certification électronique

Les informations à déclarer sur un certificat électronique sont strictement identiques à celles déclarées sur un certificat papier, à travers un site internet sur lequel l'authentification est sécurisée (www.certdc.inserm.fr).

Pour les décès certifiés électroniquement, la partie inférieure du certificat de décès, contenant les informations d'état civil non nominatives et les causes de décès, est directement transmise au CépiDc. Une application de restitution permet aux médecins de santé publique des ARS et à l'InVS d'accéder à ces informations dans leur champ territorial de responsabilité pour exercer leur mission de veille et d'alerte sanitaire. Les informations d'état civil, nominatives mais sans les causes de décès, sont transmises directement à l'INSEE.

Dans la base du CépiDc, le mode de certification (papier ou électronique) de chaque certificat est conservé. Outre la lisibilité accrue des certificats, l'utilisation de la certification électronique a tendance à améliorer la quantité et la cohérence de l'information déclarée [4].

A la date de rédaction de cet article, la partie administrative nominative sans les causes de décès doit être imprimée par le médecin certificateur et transmise à la commune de décès pour que celle-ci procède à la mise à jour de son registre d'état civil. La mairie de décès produit ensuite le bulletin 7 et

l'avis 7 bis qui suivront des parcours identiques à la procédure papier. La nécessité de cette impression et de cet envoi devrait disparaître dans les prochains mois avec les évolutions techniques et réglementaires adéquates.

Le codage médical des certificats de décès

Le codage médical des certificats de décès comporte deux activités distinctes et successives :

- attribuer un code à chaque maladie, traumatisme ou cause externe de décès mentionné sur le certificat,
- sélectionner et coder la cause initiale de décès.

Le codage détermine largement la qualité et la comparabilité internationale des données produites. La Classification internationale des maladies (CIM) définit les codes, les règles et les directives permettant de mener ces tâches à bien. La 10^{ème} révision de la CIM est utilisée au CépiDc depuis 2000.

La cause initiale de décès est définie par l'OMS comme « a) la maladie ou le traumatisme qui a déclenché l'évolution morbide conduisant directement au décès, ou b) les circonstances de l'accident ou de la violence qui ont entraîné le traumatisme mortel ». La cause initiale est donc la cause sur laquelle il faut agir pour prévenir le décès. Bien que toutes les entités nosologiques mentionnées sur le certificat soient conservées, cette cause est la principale utilisée pour présenter les statistiques de mortalité par cause.

Plusieurs règles de la CIM permettent de sélectionner et éventuellement de modifier la cause initiale du décès.

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

Les règles de modification de la cause initiale permettent de concentrer le plus d'information possible sur la cause initiale. Il est par exemple possible de distinguer un diabète à l'origine d'un coma (code E140) d'un diabète avec acidocétose (code E141).

Le travail des codeurs en mortalité (nosologistes) nécessite une connaissance parfaite de la CIM ainsi qu'une grande expérience des maladies et de leur enchaînement causal mentionné sur les certificats de décès. On observe cependant des différences de codage importantes entre les codeurs, que ce soit au niveau national ou entre les pays. C'est pour cette raison que de plus en plus de pays optent pour les systèmes de codage automatique. Ces systèmes facilitent beaucoup la mise en œuvre de la CIM dont les règles de codage se complexifient régulièrement. Le codage automatique ne supprime pas pour autant le travail des codeurs pour traiter les cas complexes. Dans ce contexte, le CépiDc de l'Inserm a développé un logiciel de codage automatique compatible avec la CIM et avec les systèmes de codage utilisés par la plupart des autres pays.

Le logiciel Iris, en vigueur pour les décès à partir de 2011, remplace le logiciel Styx utilisé pour les décès survenus de 2000 à 2010. Styx a été développé par le CépiDc [5], et Iris est le fruit d'une coopération avec Statistics Sweden et le NCHS-CDC Américain[6]. Il a été créé pour être utilisable par l'ensemble des pays, et est actuellement utilisé par une vingtaine de pays. Le déploiement de ce logiciel constitue la meilleure garantie de comparabilité internationale des bases de données de mortalité par cause. Passée la période de prise en main du logiciel, générant des sauts de tendance pour certaines causes de décès, les différentiels de choix de la cause initiale de décès entre pays sont amenés à leur niveau minimal. De plus, le logiciel documente systématiquement les choix de codage médical faits, offrant ainsi une meilleure compréhension de la base à l'utilisateur.

Iris code actuellement automatiquement 50% des décès. Les autres décès sont rejetés et revus manuellement par des codeurs nosologistes. Les rejets concernent les certificats avec des causes initiales incohérentes ou vraisemblablement erronées et les certificats qui ont trait aux enfants de moins de 28 jours et aux morts maternelles. Les certificats sans cause ou avec des causes mal définies sont également signalés afin d'effectuer par courrier des demandes d'informations complémentaires auprès des médecins certificateurs (en identifiant le médecin ou le service à l'aide du cachet apposé à la fin du certificat de décès et en indiquant les éléments connus (date de naissance, sexe, date de décès ...) pour permettre au médecin l'identification du défunt).

Modalités d'accès aux données

Les modalités d'accès aux données peuvent varier en fonction du type d'analyse :

- Pour un certain nombre d'indicateurs (effectifs, taux bruts, taux standardisés), par département, année, sexe et classe d'âges, elles sont librement accessibles (en "Open Data") sur le site du CépiDc (www.cepidc.inserm.fr) dans la mesure où les données ne permettent pas de réidentifier des sujets décédés pour en obtenir la cause de décès.
- Lorsque les demandes de données ne sont pas traitées par le site internet et que les données demandées ne sont pas indirectement identifiantes, il est possible de faire une demande plus ciblée à l'adresse diffusion.cepidc@inserm.fr. L'extraction peut être faite dans un délai de deux semaines.
- Lorsque les demandes nécessitent des données très précises et susceptibles de permettre une divulgation de l'information médicale (à l'échelle de la commune de domicile notamment) ou un chaînage individuel, une autorisation de la CNIL est requise, précédée d'un avis du CCTIRS (Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé). Ces deux étapes sont généralement les plus longues dans le processus d'obtention des données. L'extraction peut être faite dans un délai d'un mois.
- Lorsqu'un appariement avec une autre base de données est nécessaire, en l'absence d'identifiant direct commun à d'autres bases de données, les chainages entre la base des causes de décès et d'autres données se font à l'aide de variables administratives communes aux deux bases permettant une identification indirectes. Dans le cas le plus courant, les porteurs d'étude doivent suivre le protocole décrit par le décret 98-37, après demande d'autorisation à la CNIL et avis du CCTIRS : un premier chaînage est réalisé par l'INSEE avec le RNIPP pour récupérer la date et la commune de décès, et, le cas échéant, les causes de décès sont recherchées dans un second temps par chaînage à l'aide des dates de naissance et de décès, de la commune de décès et du sexe. Le plateau informatique du Centre d'Epidémiologie et de Santé des Populations (CESP) de l'Inserm agit en tant que tiers de confiance pour assurer les échanges entre le porteur de projet, l'INSEE et le CépiDc. L'ensemble de la procédure d'enrichissement prend généralement un mois.

Principaux types d'analyse sur les données

Les données de mortalité permettent d'effectuer des analyses variées sur un grand nombre de sujets relatifs à la santé publique. L'exhaustivité des données, à raison d'environ 550 000 décès par an, permet de traiter de sujets y compris sur des maladies ou évènements rares pour lesquelles peu de données sont par ailleurs exploitables. Pour des problématiques traitant de variation à court terme de la mortalité, il est possible d'utiliser des données d'état civil sans les causes de décès, produites

plus rapidement par l'INSEE (80% des données reçues dans les 10 jours suivants le décès), ou des données issues de la certification électronique produites par l'Inserm avant codage, en utilisant le texte brut saisi par le médecin (10% des décès reçu dans les 24 heures). Autrement, les données codées finales, après retour éventuel au médecin et réception des résultats d'enquêtes médico-légales, sont produites jusqu'à 24 mois après la date de décès. Nous décrivons ci-dessous uniquement les analyses faites sur données finalisées.

On peut regrouper ces types d'analyses en trois grandes catégories : les analyses sur caractéristiques obtenues à partir des certificats de décès, les analyses écologiques et les analyses longitudinales sur données chaînées à d'autres données.

Etudes sur les caractéristiques individuelles des décès

L'analyse des caractéristiques présentées sur le certificat doit être faite avec prudence, et sans oublier qu'il s'agit-là d'informations recueillies après le décès, qui peuvent par conséquent ne pas être comparables avec les informations recueillies du vivant des personnes. Il faut également bien prendre en compte que les causes de décès mentionnées sur le certificat ne constituent pas un tableau clinique de la personne décédée, mais bien le processus morbide identifié par le médecin comme ayant mené au décès, avec parfois des informations incomplètes. Par exemple, en France, parmi les décès de personnes prenant un traitement antidiabétique, seul 35% des certificats de décès mentionnent un diabète[7]. De façon très descriptive, l'analyse des certificats de décès permet de mesurer la part respective des différentes causes de décès dans la mortalité générale. Elle permet notamment de constater que la France possède une mortalité par maladies cardiovasculaires particulièrement faible, avec un nombre de décès inférieur au nombre de décès par cancer depuis 2004 [8]. Elle permet également de calculer des taux de mortalité par cause, en considérant la cause initiale uniquement ou plusieurs causes mentionnées sur le certificat. Dans ce dernier cas, on parle alors d'analyse en causes multiples.

A travers la sélection de causes particulièrement associées à un facteur de risque, il est possible d'étudier les variations, notamment temporelles et spatiales, de ses conséquences sur la mortalité. C'est notamment le cas lorsque l'on considère la mortalité dite "évitable liée aux comportements à risque", qui regroupe les causes de décès liées à la consommation d'alcool et de tabac (cancer des VADS, cancer du poumon, cirrhose, psychose alcoolique et alcoolisme), aux accidents de la route et aux pratiques sexuelles à risque (Sida et infections par le VIH), pour lesquelles la France est dans une situation particulièrement mauvaise en Europe de l'Ouest [9].

L'analyse dans le temps et entre les pays des variations du niveau de mortalité pour certaines causes a également été utilisée pour évaluer la qualité des systèmes de soin [10]. En suivant ce type

d'approche, la comparabilité dans le temps et dans l'espace des données entre les pays doit garantir que les différentiels de mortalité mesurés ne sont pas en réalité des différentiels d'évolution de codage, auquel cas des approches correctives doivent être appliquées [11].

Il est également possible de faire des analyses plus poussées, parmi lesquelles :

- L'analyse de la mortalité selon le pays de naissance [12], dont la déclaration varie a priori peu après le décès du sujet,
- L'analyse des associations entre les causes mentionnées sur le certificat [13], qui permet notamment de décrire la mortalité par hémophilie [14], ou par lupus [15], et d'identifier des causes particulièrement associées aux risques liés à l'anesthésie [16],
- L'utilisation d'une information spécifique, la mention "sans domicile fixe", pour mesurer la répartition géographique et par cause des décès de cette population [17].

Analyses écologiques

Les analyses écologiques visent à décrire la variation du risque de mortalité en fonction de caractéristiques observables à l'échelle d'un groupe de population (zone géographique, période). Le plus souvent, les analyses faites à l'échelle écologique mesurent des associations qui ne peuvent pas être directement extrapolées à une échelle individuelle. Elles sont le plus souvent considérées comme d'un faible niveau de preuve, mais peuvent néanmoins générer des hypothèses importantes pour entreprendre des analyses ultérieures. Des exemples d'analyses écologiques menées à travers les données de mortalité ont permis de mesurer l'association entre :

- Température, pollution et mortalité [18], et plus particulièrement en été et pendant les vagues de chaleur[19,20],
- Taux de chômage et taux de suicide à l'échelle des pays, particulièrement en mesurant l'effet surajouté de la crise de 2008 [21],
- Niveau socio-économique de résidence, à l'échelle disponible la plus fine de la commune, et mortalité par cause[22].

Ces données écologiques pourraient à terme être également exploitées pour traiter des problématiques de veilles et d'alerte sanitaire, en identifiant la concomitance de regroupements inattendus de décès en un lieu et sur une période donnée avec des expositions environnementales suspectées, en particulier lorsque la certification électronique des décès sera déployée de façon plus massive.

Analyse de données chaînées aux causes de décès

L'analyse longitudinale de données chaînées permet d'exploiter avec la plus grande pertinence les données des certificats de décès pour en déduire des mesures d'associations utiles à la compréhension des mécanismes et à la prise de décision en santé publique. En effet, ces données chaînées ne souffrent pas de la limite importante que constitue le recueil après décès, généralement génératrice de biais. C'est notamment le cas de la profession, dont la déclaration après décès par les proches génère des biais systématiques limitant fortement son exploitation [23].

Ces données chaînées permettent ensuite d'effectuer un nombre très divers d'exploitations comme :

- La mesure de l'association entre niveau d'étude [24], profession [25] ou trajectoires professionnelles [26] et mortalité par cause à l'aide de données de panels représentatifs de la population française ou de la population salariée,
- La mise en évidence du lien entre exposition professionnelle et mortalité dans des cohortes de travailleurs, par exemple les mineurs d'uranium [27] ou les producteurs d'acier[28],
- L'étude de l'association entre obésité, consommation d'alcool et mortalité par cause dans de larges cohortes internationales [29,30],
- La mesure des différentiels de mortalité post-hospitalière selon l'établissement d'hospitalisation [31], et les très nombreuses études possibles à travers la base appariée entre le SNIIRAM-PMSI (Système d'information inter-régimes de l'assurance maladie, déjà chaîné au données hospitalières) et les causes de décès.

Discussion et perspectives

Le présent article illustre à quel point une méthodologie relativement simple de production des données permet de produire des résultats sur un grand nombre de problématiques de santé publique.

Qualité et comparabilité des données

L'ensemble des analyses faites à partir des données de mortalité par cause est conditionnée par la qualité et le niveau de comparabilité des données selon les catégories étudiées (pays, période, lieu de décès...). C'est pourquoi l'évaluation de la qualité des données est une question récurrente permettant de quantifier, et parfois de corriger certains biais [32]. Au niveau européen, à la fin des années 1990, une large étude a fait un état des lieux des différents systèmes de production [33]. Elle mettait en évidence qu'alors que la certification des décès doit obligatoirement être faite par un médecin dans la plupart des pays, plusieurs sources de non comparabilité des données pouvaient

être identifiées : les certificats de décès utilisés dans les différents pays n'étaient pas toujours conformes au format recommandé par l'OMS, la notice fournie avec le certificat pour aider le médecin à le remplir n'était pas identique, la formation des médecins à la certification était insuffisante et très disparate et les procédures de recoupement avec des données d'état civil n'étaient pas systématiques. Un aspect relativement important peut affecter la comparabilité des données : les causes de décès sont strictement confidentielles dans une large partie de l'Europe (Allemagne, Pays-Bas, Autriche, Belgique, France, Espagne, Italie, Luxembourg) et sont publiques ou communicables pour d'autres pays (Europe du Nord, Portugal, Royaume-Uni). Par ailleurs, les procédures d'investigation des morts suspectes sont variables, avec des enquêtes médico-légales non systématiques dans certains pays dont la France, et des procédures plus approfondies, notamment au Royaume-Uni, où deux coroners doivent certifier les décès suite à des autopsies systématiques. Ces éléments affectent de façon substantielle la comparabilité des données [34], notamment de mortalité par suicide [35]. Plusieurs démarches permettent d'apporter des éléments d'amélioration de la qualité et de la comparabilité des données, que ce soit à l'étape de la déclaration du médecin, du processus de codage ou de l'exploitation statistique.

La première démarche est l'amélioration de la qualité de la certification par le médecin. Une première voie pour y parvenir consiste à mettre en place les conditions pour que le médecin certificateur d'un décès connaisse le sujet décédé (médecin traitant), même lorsque le décès survient à domicile. Une seconde voie est la mise en place et l'utilisation de la certification électronique qui permet d'apporter un support en ligne d'aide à la certification, d'améliorer la sécurité et la qualité du traitement de l'information, de limiter son coût et la recopie manuelle de l'information administrative sur différents supports, et surtout d'accélérer très fortement la procédure, passant d'environ 3 mois pour le support papier à une transmission quasi-immédiate pour l'électronique. La montée en charge de l'utilisation de l'application est toutefois encore aujourd'hui assez lente, notamment en raison des nombreuses démarches administratives encadrant un décès encore souvent fondées sur l'émission d'un papier. Dans le cadre de la politique générale actuelle de modernisation de l'état et de simplification des procédures, un projet interministériel est à l'étude pour dématérialiser intégralement la déclaration des décès, permettant également pour un médecin d'effectuer la certification sur support mobile, possibilité fort utile lorsque le décès ne survient pas à l'hôpital. A terme, l'objectif affiché est d'atteindre 100% de décès certifiés électroniquement, objectif qui pour être atteint nécessitera que les différentes institutions concernées se mobilisent, comme cela a été le cas au Portugal où cet objectif est atteint depuis 2014. Parallèlement, le système de certification électronique pourrait également constituer un moyen d'interagir avec le médecin pour

lui faciliter la saisie de la partie médicale du certificat, et éventuellement lui signaler une possible incohérence dans sa déclaration.

La seconde démarche consiste à continuer de promouvoir l'extension de l'utilisation du logiciel Iris à l'international. Aujourd'hui officiellement utilisé par 19 pays essentiellement européens, le logiciel est également en phase de test dans de nombreux autres pays couvrant l'ensemble des continents et de nombreuses langues. Au-delà d'une automatisation de plus en plus avancée des règles de codage indépendante du langage utilisées pour choisir la cause initiale de décès, une perspective d'amélioration du logiciel consiste à intégrer de façon croissante des éléments de traitement automatisé du langage naturel. Cette intégration viserait à mettre à profit tous les développements récents réalisés dans ce domaine avec le développement d'internet, aussi bien sur le plan syntaxique que sémantique. Ce type de développement est indispensable pour faciliter l'exploitation en temps quasi-réel de l'information pour la veille et l'alerte sanitaire, il peut et doit également être un moyen d'améliorer encore la comparabilité internationale des données, en visant à ce que les méthodes retenues puissent être reproductibles et adaptées à différentes langues.

Une troisième démarche consiste à favoriser l'exploitation des données en causes multiples, perspective impliquant une évolution du processus de production. En effet, à ce jour, seule la cause initiale de décès fait l'objet d'une définition formelle. A ce titre, elle est la seule à pouvoir être facilement exploitée pour des comparaisons internationales. Pourtant, de nombreux pays conservent toutes les causes de décès mentionnées sur les certificats, qui deviennent de plus en plus nombreuses dans le contexte du vieillissement de la population et de la multiplication des maladies chroniques qui l'accompagne. En moyenne, 3,4 causes sont déclarées par certificat en France, et la pertinence du concept de cause initiale, et de la démarche de certification dans ce contexte peuvent être remis en question [36]. C'est pourquoi les initiatives pour exploiter la totalité des causes mentionnées sur les certificats sont de plus en plus nombreuses. Une des approches déjà largement employée est l'étude des associations entre causes [37,38]. Toutefois, l'utilisation de cette information se limite généralement à décompter la présence de causes sur le certificat, indépendamment de leur position [13,39]. Cette méthode, appelée "approche en causes multiples", permet notamment de trouver de nombreux décès liés à des pathologies chroniques comme le diabète ou l'hypertension parallèlement à leur décompte en cause initiale [40]. L'approche en cause multiple actuelle pose cependant des problèmes d'interprétations. En effet, il n'est pas simple d'attribuer un statut aux causes de décès décomptées, et d'évaluer l'importance de leur rôle joué dans le processus létal. En outre, dénombrer toutes les causes indépendamment de toutes autres considérations revient à ce que les décès pour lesquels plus de causes sont mentionnées aient un

poids plus grand dans une analyse en causes multiples. Une approche plus pertinente doit être développée.

Système national des données de santé

Dans le contexte de la loi de modernisation du système de santé, il est prévu la création d'un Système national des données de santé (SNDS). Ce système viserait à centraliser et à chaîner de façon pérenne notamment les bases de données du SNIIRAM, du PMSI et des causes médicales de décès, toutes les trois ayant la particularité de couvrir la quasi-totalité de la population française.

A travers ces données, il deviendra donc notamment possible de mesurer l'association entre tout évènement (acte médical, achat de médicaments, hospitalisation) ayant fait l'objet d'un remboursement par l'assurance maladie et les causes de décès. Ceci permettra un nombre potentiellement considérable d'applications importantes pour mesurer l'effet bénéfique ou délétère à court, moyen ou long terme, de pratiques médicales sur la mortalité par cause, enrichissant largement les études de pharmacoépidémiologie déjà menées sur le SNIIRAM-PMSI [41]. En utilisant des algorithmes actuellement en fort développement, il est par ailleurs possible de repérer de nombreuses pathologies à travers les consommations de soin observées. On peut donc présupposer que cette base de données constituera une source importante de connaissance pour suivre le devenir de patients atteints de maladies spécifiques.

Plus particulièrement, un enrichissement de la base des causes médicales de décès par des informations issues du SNDS à l'étape du codage médical par le CépiDc serait à étudier. Une telle mise en regard de l'information permettrait notamment de préciser des causes de décès, voire de compléter des informations manquantes par des informations sur la prise de médicaments (SNIIRAM) ou des diagnostics hospitaliers (PMSI). Des questions méthodologiques importantes se poseront alors pour exploiter ces données de façon optimale, en particulier pour veiller à ce que cela ne crée pas artificiellement des associations entre prises de médicament et causes de décès.

Conclusion

Comme dans la plupart des pays industrialisés, la production de la base des causes médicales de décès en France est fondée sur un acte médical obligatoire et un protocole de remontée d'information relativement stable dans le temps et dans l'espace. Ses caractéristiques permettent à de nombreux acteurs d'en faire une exploitation essentielle pour traiter de nombreuses problématiques de santé publique d'observation, de veille sanitaire, et de recherche.

Pour améliorer la qualité et la comparabilité des statistiques issues de ces données, plusieurs pistes sont envisagées et en cours de mise en œuvre, qui devraient encore accroître leur utilisation. Elles nécessiteront la mobilisation coordonnée des acteurs allant du médecin certificateur, au statisticien exploitant les données.

Références

- [1] Blayo Y. La mortalité en France de 1740 à 1829. *Population* 1975;30:123–42.
- [2] Vallin J, Meslé F. Origine et histoire de la statistique française des causes de décès. *Causes Décès En Fr. 1925 À 1978*. INED, Paris: n.d., p. 9–89.
- [3] Loi relative à la politique de santé publique. 2004.
- [4] Lefeuvre D, Aouba A, Pavillon G, Lamarche-Vadel A, Fouillet A, Jouglu E, et al. Quality comparison of electronic versus paper death certificates (France – 2010). *Popul Health Metr* 2014.
- [5] Pavillon G, Jouglu E. The French automated coding system Styx. *Proc. Int. Collab. Effort Autom. Mortal. Stat.*, vol. 2, Bethesda, Maryland, USA: 2001, p. 50–1.
- [6] Pavillon G, Johansson LA, Weber S. Iris – International automated coding system for causes of death. 2012.
- [7] Romon I, Rey G, Mandereau-Bruno L, Weill A, Jouglu E, Eschwège E, et al. The excess mortality related to cardiovascular diseases and cancer among adults pharmacologically treated for diabetes--the 2001-2006 ENTRED cohort. *Diabet Med J Br Diabet Assoc* 2014;31:946–53. doi:10.1111/dme.12435.
- [8] Aouba A, Jouglu E, Eb M, Rey G. L'évolution de la mortalité et des causes de décès entre 1990 et 2009. *Actual Doss Santé Publique* 2012:24–8.
- [9] Jouglu E. Indicateurs de mortalité “prématurée” et “évitable.” Paris: Haut Conseil de la Santé Publique; 2013.
- [10] Mackenbach JP, Hoffmann R, Khoshaba B, Plug I, Rey G, Westerling R, et al. Using “amenable mortality” as indicator of healthcare effectiveness in international comparisons: results of a validation study. *J Epidemiol Community Health* 2013;67:139–46. doi:10.1136/jech-2012-201471.
- [11] Rey G, Aouba A, Pavillon G, Hoffmann R, Plug I, Westerling R, et al. Cause-specific mortality time series analysis: a general method to detect and correct for abrupt data production changes. *Popul Health Metr* 2011;9:52. doi:10.1186/1478-7954-9-52.
- [12] Boulogne R, Jouglu E, Breem Y, Kunst AE, Rey G. Mortality differences between the foreign-born and locally-born population in France (2004-2007). *Soc Sci Med* 1982 2012;74:1213–23. doi:10.1016/j.socscimed.2012.01.002.
- [13] Desesquelles A, Salvatore MA, Frova L, Pace M, Pappagallo M, Meslé F, et al. Revisiting the mortality of France and Italy with the multiple-cause-of-death approach. *Demogr Res* 2010;23:71–806.
- [14] Aouba A, Rey G, Pavillon G, Jouglu E, Rothschild C, Torchet M-F, et al. Deaths associated with acquired haemophilia in France from 2000 to 2009: multiple cause analysis for best care strategies. *Haemoph Off J World Fed Hemoph* 2012;18:339–44. doi:10.1111/j.1365-2516.2011.02647.x.
- [15] Thomas G, Mancini J, Jourde-Chiche N, Sarlon G, Amoura Z, Harlé J-R, et al. Mortality Associated With Systemic Lupus Erythematosus in France Assessed by Multiple-Cause-of-Death Analysis. *Arthritis Rheumatol* 2014;66:2503–11. doi:10.1002/art.38731.
- [16] de Saint Maurice G, Aouba A, Pequignot F, Auroy Y, Benhamou D, Jouglu E, et al. Anesthesia-related mortality. *Anesthesiology* 2009;111:1165; author reply 1166. doi:10.1097/ALN.0b013e3181bbc49b.
- [17] Vuillermoz C, Aouba A, Grout L, Vandentorren S, Tassin F, Vazifeh L, et al. Estimating the number of homeless deaths in France, 2008-2010. *BMC Public Health* 2014;14:690. doi:10.1186/1471-2458-14-690.
- [18] Todd N, Valleron A-J. Space-Time Covariation of Mortality with Temperature: A Systematic Study of Deaths in France, 1968-2009. *Environ Health Perspect* 2015;123:659–64. doi:10.1289/ehp.1307771.

- [19] Fouillet A, Rey G, Jouglu E, Frayssinet P, Bessemoulin P, Hémon D. A predictive model relating daily fluctuations in summer temperatures and mortality rates. *BMC Public Health* 2007;7:114. doi:10.1186/1471-2458-7-114.
- [20] Analitis A, Michelozzi P, D'Ippoliti D, Donato F De', Menne B, Matthies F, et al. Effects of heat waves on mortality: effect modification and confounding by air pollutants. *Epidemiol Camb Mass* 2014;25:15–22. doi:10.1097/EDE.0b013e31828ac01b.
- [21] Laanani M, Ghosn W, Jouglu E, Rey G. Impact of unemployment variations on suicide mortality in Western European countries (2000-2010). *J Epidemiol Community Health* 2015;69:103–9. doi:10.1136/jech-2013-203624.
- [22] Rey G, Jouglu E, Fouillet A, Hemon D. Ecological association between a deprivation index and mortality in France over the period 1997 - 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health* 2009;9:33.
- [23] Rey G, Rican S, Luce D, Menvielle G, Jouglu E. Measuring social inequalities in cause-specific mortality in France: comparison between linked and unlinked approaches. *Rev Dépidémiologie Santé Publique* 2013;61:221–31. doi:10.1016/j.respe.2012.11.004.
- [24] Menvielle G, Chastang JF, Luce D, Leclerc A. [Changing social disparities and mortality in France (1968-1996): cause of death analysis by educational level]. *Rev Epidemiol Sante Publique* 2007;55:97–105.
- [25] Cohidon C, Santin G, Geoffroy-Perez B, Imbernon E. [Suicide and occupation in France]. *Rev Epidemiol Sante Publique* 2010;58:139–50.
- [26] Karimi M, Geoffroy-Perez B, Fouquet A, Latouche A, Rey G. Socioprofessional trajectories and mortality in France, 1976-2002: a longitudinal follow-up of administrative data. *J Epidemiol Community Health* 2014. doi:10.1136/jech-2014-204615.
- [27] Rage E, Caër-Lorho S, Drubay D, Ancelet S, Laroche P, Laurier D. Mortality analyses in the updated French cohort of uranium miners (1946-2007). *Int Arch Occup Environ Health* 2015;88:717–30. doi:10.1007/s00420-014-0998-6.
- [28] Bourgkard E, Wild P, Courcot B, Diss M, Ettliger J, Goutet P, et al. Lung cancer mortality and iron oxide exposure in a French steel-producing factory. *Occup Environ Med* 2009;66:175–81. doi:10.1136/oem.2007.038299.
- [29] Bergmann MM, Rehm J, Klipstein-Grobusch K, Boeing H, Schütze M, Drogan D, et al. The association of pattern of lifetime alcohol use and cause of death in the European prospective investigation into cancer and nutrition (EPIC) study. *Int J Epidemiol* 2013;42:1772–90. doi:10.1093/ije/dyt154.
- [30] Lajous M, Bijon A, Fagherazzi G, Boutron-Ruault M-C, Balkau B, Clavel-Chapelon F, et al. Body mass index, diabetes, and mortality in French women: explaining away a “paradox.” *Epidemiol Camb Mass* 2014;25:10–4. doi:10.1097/EDE.0000000000000031.
- [31] Lamarche-Vadel A, Ngantcha M, Le Pogam M-A, Ghosn W, Grenier C, Meyer L, et al. Hospital comparisons based on mortality: revisiting the choice of post-admission timeframe and evaluating the contribution of the causes of death, France, 2009. *Med Care* 2015.
- [32] Phillips DE, Lozano R, Naghavi M, Atkinson C, Gonzalez-Medina D, Mikkelsen L, et al. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. *Popul Health Metr* 2014;12:14. doi:10.1186/1478-7954-12-14.
- [33] Jouglu E, Rossollin F, Niyonsenga A, Chappert J-L, Johansson LA, Pavillon G. Quality and comparability improvement of European causes of death statistics. 2001.
- [34] Jouglu E, Salem G, Gancel S, Michel V, Kürzinger M-L, Pavillon G, et al. Atlas de la mortalité dans l'union européenne. Données 1994-1996. Eurostat; 2002.
- [35] Aouba A, Péquignot F, Camelin L, Jouglu E. Évaluation de la qualité et amélioration de la connaissance des données de mortalité par suicide en France métropolitaine, 2006. *Bul Epidémiol Hebd* 2011:497–500.
- [36] Alpérovitch A, Bertrand M, Jouglu E, Vidal J-S, Ducimetière P, Helmer C, et al. Do we really know the cause of death of the very old? Comparison between official mortality statistics and cohort study classification. *Eur J Epidemiol* 2009;24:669–75. doi:10.1007/s10654-009-9383-2.

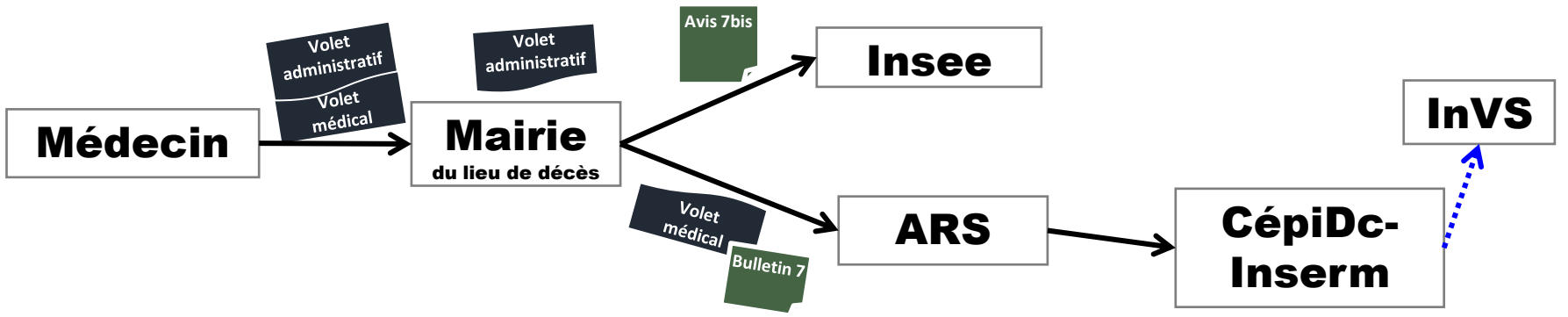
- [37] Viallon V, Banerjee O, Jouglu E, Rey G, Coste J. An empirical comparison study of approximate methods for structure selection in binary graphical models. *Biom J* 2014.
- [38] Redelings MD, Wise M, Sorvillo F. Using Multiple Cause-of-Death Data to Investigate Associations and Causality between Conditions Listed on the Death Certificate. *Am J Epidemiol* 2007;166:104–8. doi:10.1093/aje/kwm037.
- [39] Redelings MD, Sorvillo F, Simon P. A comparison of underlying cause and multiple causes of death: US vital statistics, 2000–2001. *Epidemiol Camb Mass* 2006;17:100–3.
- [40] Romon I, Jouglu E, Balkau B, Fagot-Campagna A. The burden of diabetes-related mortality in France in 2002: an analysis using both underlying and multiple causes of death. *Eur J Epidemiol* 2008;23:327–34. doi:10.1007/s10654-008-9235-5.
- [41] Martin-Latry K, Bégau B. Pharmacoepidemiological research using French reimbursement databases: yes we can! *Pharmacoepidemiol Drug Saf* 2010;19:256–65. doi:10.1002/pds.1912.

Figures

Figure 1 : Circuits papier et électronique de transmission des certificats de décès à l'Inserm

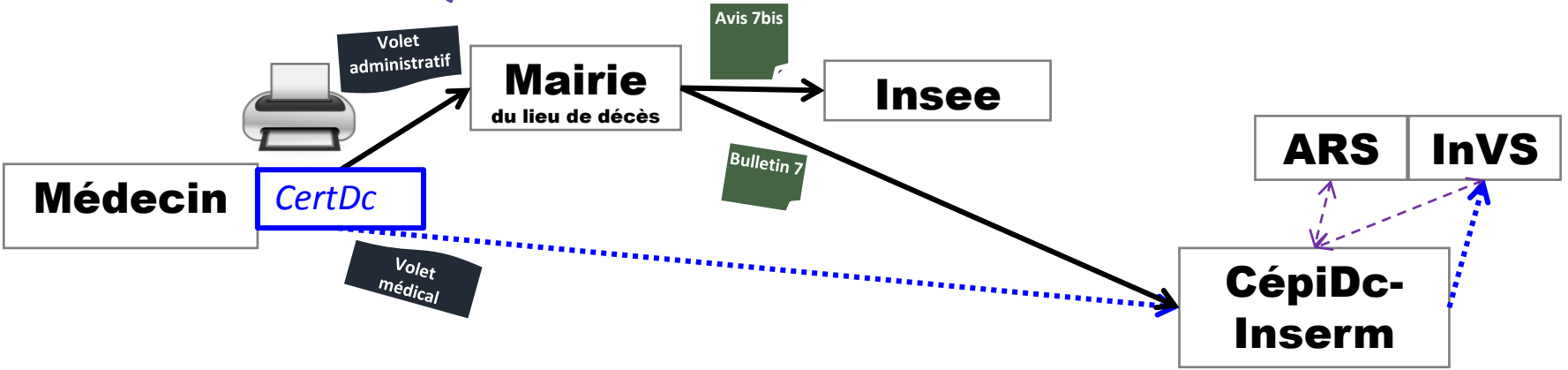
Figure 2 : Certificat médical de décès

Transmission PAPIER



Disponibilité des données : 3 semaines à 4 mois

Transmission ELECTRONIQUE



Disponibilité des données : quasi immédiate

→ Transmission papier Flux informatique <---> Accès web

A remplir par le Médecin

COMMUNE DE DÉCÈS :

Code Postal

NOM :

Prénoms :

Date de naissance : Sexe :

Domicile :

Le docteur en médecine soussigné, certifie que la mort de la personne désignée ci-contre, survenue le _____ à _____ heure _____ est réelle et constatée (voir 1 au verso)

Obstacle médico-légal (voir 2 au verso) OUI NON

Obligation de mise en bière immédiate (voir 3 au verso) OUI NON

- dans un cercueil hermétique (voir 4 au verso) OUI NON

- dans un cercueil simple (voir 5 au verso) OUI NON

Obstacle au don du corps (voir 6 au verso) OUI NON

Prélèvement en vue de rechercher la cause du décès (voir 7 au verso) OUI NON

Présence de prothèse fonctionnant au moyen d'une pile (voir 8 au verso) OUI NON

Important : bien cacher toutes les lignes par oui ou non

A _____ le _____
Signature (Non lisible) et Cachet (obligatoire) du médecin

RÉSERVÉ À LA MAIRIE

Le numéro d'ordre du décès sur le registre des actes de l'état civil à inscrire ci-contre doit être reproduit au verso.

N° D'ORDRE du décès

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

A conserver dans la mairie du lieu d'implantation de la chambre funéraire.

A remplir et à clore par le Médecin

Renseignements confidentiels et anonymes

Code Postal : Commune de décès :

Date de décès :

1. Sexe masculin

Code Postal : Commune de domicile :

Date de naissance :

2. Sexe féminin

Causes du décès

PARTIE I Maladie(s) ou affection(s) morbide(s) ayant directement provoqué le décès *

La dernière ligne remplie doit correspondre à la cause initiale.

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

a) _____

due à ou consécutive à b) _____

due à ou consécutive à c) _____

due à ou consécutive à d) _____

* Il s'agit de la maladie, du traumatisme, de la complication ayant entraîné la mort (et non du mode de décès, ex. : syncope, arrêt cardiaque...)

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

Informations complémentaires

Le décès est-il survenu pendant une grossesse (à déclarer, même si cet état n'a pas contribué à la mort) ou moins d'un an après ? 1. Oui 2. Non

Dans ce dernier cas, intervalle entre la fin de cette grossesse et le décès : Mois Jours

En cas d'accident, préciser le lieu exact de survenue (voie publique, domicile...):

S'agit-il d'un accident du travail (ou présumé tel) ?

1. Oui 2. Non 3. Sans précision

Autopsie : une autopsie a-t-elle été ou sera-t-elle pratiquée ?

Lieu du décès :

1. Non 2. Oui, résultat disponible

1. Domicile

2. Hôpital

3. Clinique privée

3. Oui, résultat non disponible

4. Hospice, maison de retraite

5. Voie publique

6. Autre lieu

Signature (Non lisible) et Cachet (obligatoire) du médecin

Exemples

I. a) Septicémie b) Périlésion c) Prééclampsie d'obère d) Ulcère duodinal II. Alcoolisme	Intervalle 1 à 4 18 à 21 1 à 7 7	I. a) Cerveau b) Oedème cérébral c) Infection crânienne d) Accident de la route II. Tumeur	Intervalle 22 à 28 à 21 21	I. a) Hémostase cérébrale b) Hypertension c) d) II. Cancer du sein récidivé	Intervalle 1 à 18 à 18 à 7
I. a) Embolisme cérébral b) SIDA c) d) II. Kaposi, Tuberculose	Intervalle 20 à 7 à - -	I. a) Noyau b) Saisie c) d) II. Tumeur	Intervalle -	I. a) Défaillance respiratoire b) Embolie pulmonaire c) Pneuonie d) Accouchement II. Virus	Intervalle 1 à 3 à 7 à 16 à -