

Direction de la recherche, des études,
de l'évaluation et des statistiques
DREES

SÉRIE
ÉTUDES

**DOCUMENT
DE
TRAVAIL**

Élaboration d'un outil de mesure
de la mortalité post hospitalière
Tome 1 - Étude de faisabilité

Olivier VALLET, Sylvie VILLEMINOT,
Isabelle GASQUET, Bruno FALISSARD
Hôpital Paul Brousse – Département de santé publique
n° 40 – janvier 2004

SOMMAIRE

A. Introduction	7
B. Objectifs	7
C. Accès aux données de mortalité : la situation en France	8
C.1 Circuit des différents documents relatifs a un décès.....	8
C.1.1 Traitement de l'avis de décès par l'INSEE	9
C.1.2 Traitement des certificats médicaux de causes de décès par l'INSERM	10
C.2 Recherche du statut vital des personnes dans le domaine médical	13
C.2.1 Recherche du statut vital à partir d'informations nominatives	13
C.2.2 Recherche du statut vital en l'absence d'informations nominatives.....	14
D. Systèmes de chaînage à l'étranger	15
D.1 Grande-Bretagne.....	16
D.1.1 Écosse	16
D.1.2 Angleterre	18
D.1.3 Région d'Oxford	19
D.2 Canada.....	20
D.2.1 Generalized Record Linkage System.....	20
D.2.2 Système de Manitoba.....	22
D.3 Australie occidentale	22
D.4 Suède	25
D.5 Finlande	27
D.6 États-Unis	27
E. Systèmes d'anonymisation et de chaînage : la situation en France	30
E.1 Anonymisation.....	30
E.1.1 Rappel réglementaire	30
E.1.2 Techniques d'anonymisation	32
E.2 Techniques de chaînage	33
E.3 Appariements sécurisés : exemples d'applications en France	34
E.3.1 DIM du CHU de Dijon.....	36
E.3.2 CESSI/CNAMTS	37
E.3.3 Autres dispositifs.....	44
F. La mortalité comme indicateur d'évaluation de la qualité des soins	48
F.1 Contraintes méthodologiques	48
F.2 Qualité des données.....	49
F.3 Définition de la période d'observation	50
F.4 Conclusion	51
F.5 Impact de la publication d'un indicateur de mortalité	51
G. Conclusion et perspectives	54
H. Références	56

Résumé

Le décès constitue un évènement objectif majeur dont la prise en compte dans l'évaluation des prises en charge, notamment hospitalières, est incontournable.

Actuellement en France les données de morbidité hospitalière ne peuvent être reliées qu'à la seule mortalité intra-hospitalière, recueillie dans le cadre du Pmsi. Hors la mortalité intra-hospitalière dépend fortement de facteurs indépendants de la qualité des soins, notamment de la politique d'admission et de sortie de l'établissement et de la durée de séjour. Afin d'améliorer la qualité d'un « indicateur mortalité », il serait souhaitable d'éviter ces facteurs de confusion en utilisant des données de mortalité globale, intra et extrahospitalière. Pour mettre en place un tel indicateur, il faudrait pouvoir apparier des données hospitalières individualisées à une base de mortalité.

Plusieurs pays disposent de systèmes d'information leur permettant d'accéder ainsi aux données de mortalité post-hospitalière. Les applications de ces systèmes varient selon les pays : études épidémiologiques, recherche clinique, recherche en santé publique, économie de la santé, évaluation des pratiques ou démarches qualité (indicateurs de résultats, benchmarking).

Actuellement en France, la recherche du statut vital des personnes ne peut se faire qu'à l'aide d'identifiants nominatifs alors que les données hospitalières sont disponibles uniquement sous forme anonymisée (Rsa issus du Pmsi).

Pour apparier une base de mortalité à une base hospitalière, tout en respectant la législation française concernant l'utilisation de données nominatives dans le domaine de la santé, il faut appliquer aux deux bases une même procédure d'anonymisation, créant un code anonyme identique pour un même individu. Grâce à ce code, le statut vital et les données de morbidité d'un patient pourraient être mises en relation.

Parmi les procédés d'anonymisation, la technique de hachage est actuellement préconisée par la Cnil. Cette technique consiste en une compression numérique transformant les données d'origine en un message de taille fixe. Le hachage est appliqué à des identifiants nominatifs communs aux fichiers que l'on veut apparier. Le code obtenu étant toujours le même pour un individu donné, un rapprochement des différents enregistrements le concernant peut être réalisé. Le hachage est irréversible, il est impossible de revenir aux informations d'origine à partir du code.

Plusieurs systèmes d'appariement de bases de données, préalablement anonymisées par une technique de hachage, ont déjà été développés en France. Dans le domaine hospitalier, le dispositif Fonction d'occultation d'identifiants nominatifs (FOIN) est désormais intégré au Pmsi dans tous les établissements publics et privés. Un système de double anonymisation, sécurisé par des clés secrètes, crée pour chaque patient un numéro anonyme totalement

irréversible qui permet le chaînage de tous les séjours Mco et Ssr le concernant (que ce soit dans un ou plusieurs établissements).

Des systèmes d'anonymisation et de chaînage agréés par la Cnil et d'ores et déjà opérationnels sont donc disponibles en France. Le même modèle pourrait être adopté pour apparier la base Pmsi à une base de mortalité. La faisabilité technique, institutionnelle et réglementaire d'un tel dispositif fera l'objet d'une prochaine étude.

Ce rapport se termine par une revue de la littérature scientifique sur l'intérêt et les limites de l'utilisation de la mortalité comme indicateur de qualité d'une prise en charge et sur les perspectives que la mise à disposition de la mortalité extrahospitalière ouvrirait dans d'autres champs de recherche.

A. Introduction

En France comme dans l'ensemble des pays industrialisés la contrainte budgétaire conduit à élaborer des indicateurs permettant de juger de l'efficacité du système sanitaire, non plus du seul point de vue économique (Drg – Pmsi) mais en tentant d'y associer des outils d'évaluation de la qualité de l'offre de soins, tant en termes de respect des procédures que de résultats sanitaires.

Parallèlement, une politique d'information et de prise en compte de l'avis des usagers s'est développée depuis quelques années. Déjà inscrite dans différents textes officiels (en particulier les rapports du Haut comité de santé publique), elle a été réaffirmée très récemment par la loi de mars 2002 « Droits des malades et qualité du système de santé ».

Il existe enfin une forte pression de la part des médias qui ont pris l'initiative de publier des « classements » hospitaliers, réalisés en partie à partir d'informations issues du PMSI, qui, compte tenu de la méthodologie utilisée, ne peuvent prétendre refléter des différences dans la qualité des soins.

Un travail est donc mené actuellement par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees – Ministère de la santé) pour réfléchir à l'élaboration de nouveaux indicateurs permettant d'évaluer et de comparer les performances du système de santé français ^{[26] [27]}.

Parmi les indicateurs de résultat, le décès constitue un évènement objectif majeur dont la prise en compte dans l'évaluation des prises en charge, notamment hospitalières, paraît incontournable.

Actuellement en France les données de morbidité hospitalière ne peuvent être reliées qu'à la seule mortalité intra-hospitalière, recueillie dans le cadre du Pmsi. Or cette dernière ne représente qu'une fraction de la mortalité et dépend fortement de facteurs indépendants de la qualité des soins. Son utilisation comme indicateur de résultat n'est donc pas satisfaisante et il serait souhaitable de pouvoir disposer de la mortalité post-hospitalière.

B. Objectifs

La mise en place d'un système d'information permettant d'apparier une base de morbidité hospitalière à une base de mortalité est-elle envisageable en France ? C'est à cette question que nous tenterons de répondre en étudiant la faisabilité d'un tel système, compte tenu des contraintes réglementaires et techniques en vigueur.

Ce premier document présente les résultats d'une recherche documentaire portant sur les points suivants :

- Quel est en France le circuit de l'information « décès » et comment peut-on obtenir le statut vital d'un individu ?
- Comment certains pays disposent-ils des données de mortalité extrahospitalière ? Nous décrirons quelques systèmes développés à l'étranger.
- Quelles sont les contraintes légales à respecter (en termes de confidentialité des données) ?

- Des possibilités « techniques » permettant d'apparier des bases de données anonymisées existent-elles déjà en France ?
- La mortalité est-elle un bon indicateur de qualité d'une prise en charge ?
- Quels autres champs de recherche pourraient bénéficier de la mise à disposition de données de mortalité globale anonymisées ?

C. Accès aux données de mortalité : la situation en France

C.1 Circuit des différents documents relatifs a un décès

Le **certificat de décès**, établi par le médecin qui constate le décès, est constitué de 2 parties :

⇒ La partie supérieure, destinée au bureau de l'état civil de la mairie, est **nominative**. Elle comprend des informations administratives :

- Commune de décès,
- Nom, prénoms, date de naissance, sexe,
- Adresse du domicile,
- Date et heure du décès,
- Rubriques concernant les opérations funéraires,
- Lieu, date, signature et cachet du médecin,
- Numéro d'ordre du décès (qui sera rempli par la mairie).

⇒ La partie inférieure qui constitue le Certificat médical des causes de décès (CMCD) est **anonyme**. Elle comporte des données administratives et médicales :

- code postal et commune de décès,
- code postal et commune de domicile,
- date du décès,
- date de naissance,
- sexe,
- causes du décès,
- informations complémentaires (décès pendant une grossesse ou moins d'un an après, si accident, précisions),
- autopsie,
- lieu du décès (domicile, hôpital, clinique privée, voie publique, autre lieu, sans précision),
- signature et cachet du médecin.

La partie inférieure du certificat de décès est cachetée par le médecin pour en garantir le caractère confidentiel.

Le certificat est transmis à la mairie de la commune de décès dans un délai de 24 à 48 heures.

À la mairie

L'officier d'état civil :

1. **Enregistre la déclaration de décès** à l'aide de la partie supérieure nominative du certificat (certaines données d'état civil complémentaires sont obtenues auprès de la famille ou de l'organisme funéraire et sont vérifiées sur l'acte de naissance ou le livret de famille). La partie supérieure du certificat de décès est conservée par la mairie.
2. **Établit le bulletin de décès n° 7**, non nominatif, en complétant les informations suivantes :
 - lieu (commune et département) et date du décès,
 - numéro de l'acte et numéro d'ordre de l'avis de décès,
 - sexe, date de naissance, lieu de naissance (commune et département),
 - profession, nationalité,
 - domicile (commune et département),
 - état matrimonial,
 - lieu du décès (domicile /hôpital /clinique privée /maison de retraite, hospice /voie publique /autre),
 - présence du certificat médical.
3. **Rédige l'avis de décès n° 7bis nominatif**, qui comporte, outre le nom et les prénoms, les mêmes informations que le bulletin de décès.

Le bulletin de décès n° 7 est agrafé à la partie inférieure cachetée du certificat médical de décès (au dos duquel l'officier d'état civil a noté la commune, le département, le numéro de l'acte et le numéro d'ordre du décès). Ces 2 documents, anonymes, sont transmis à la Direction départementale des affaires sanitaires et sociales (DDASS), dans un délai théorique de 7 jours. Après avoir pris connaissance des causes du décès, le médecin de la Ddass transmet tous les mois les bulletins et les certificats toujours agrafés au Centre d'épidémiologie sur les causes de décès, le CépiDc, qui traite les causes de décès.

L'avis de décès n° 7bis est adressé à l'Insee dans un délai théorique de 7 jours (sur support papier ou informatique selon les mairies).

Ce circuit est le même pour toutes les personnes décédées en France, qu'elles soient françaises ou non, résidant en France ou non.

C.1.1 Traitement de l'avis de décès par l'Insee

À partir des informations recueillies sur l'avis de décès 7bis, l'Insee procède aux opérations suivantes :

1. Radiation du Répertoire national d'identification des personnes physiques (RNIPP). Le Rnipp est géré par l'Insee pour la France et les départements d'Outre-Mer et par la Caisse nationale d'assurance vieillesse (CNAV) pour l'étranger et les Territoires d'Outre-Mer.
2. Mise à jour journalière du Répertoire national inter-régimes des bénéficiaires de l'assurance maladie (RNIAM). Nominatif, le Rniam est une copie du Rnipp complétée par le régime d'affiliation. Il concerne l'ensemble des caisses et est géré par la Caisse Nationale d'assurance vieillesse des travailleurs salariés (CNAVTS). Il est utilisé pour l'affiliation à un régime de sécurité sociale (et un seul), pour la création de cartes de

santé et l'attribution immédiate de la carte Vitale aux nouveaux-nés [décret 96-793 du 12/09/96]. L'Insee a mis en place une application, l'Alimentation informatisée du répertoire des personnes physiques (AIREPP), afin d'accélérer la mise à jour des répertoires. Les données sur les décès sont transmises par les mairies au Centre national informatique de Nantes (CNIN) dans un délai maximum théorique de 7 jours (sous forme informatique ou non). En pratique, 60 à 70 % des décès sont transmis dans la journée et l'on peut considérer qu'en 1 mois, la quasi totalité est connue.

3. Mise à jour du fichier longitudinal ou « panel » appelé Échantillon démographique permanent (EDP),
4. Mise à jour du fichier électoral,
5. Transmission aux organismes fiscaux,
6. Constitution du fichier annuel de données non nominatives des personnes décédées destiné au CépiDc.

Documentation	Document interne communiqué par l'Insee et site Internet
Contacts	Réunion avec Guy Desplanques (Insee)

C.1.2 Traitement des certificats médicaux de causes de décès par l'Inserm

La statistique des causes de décès, auparavant du ressort de l'Insee, est gérée par le Centre d'épidémiologie sur les causes médicales de décès (CépiDc, ex SC8) depuis 1968.

Le CépiDc reçoit :

- les Certificats médicaux de causes de décès (CMCD) agrafés aux bulletins de décès n° 7, envoyés mensuellement par les Ddass,
- une bande informatique comportant les données socio-démographiques des personnes décédées, envoyée une fois par an par l'Insee.

Ces documents ne comportent aucune donnée directement nominative (ni nom ni prénom).

Traitement des certificats de décès par le CépiDc:

Le traitement des causes de décès complète le fichier informatique de l'Insee, à l'aide des opérations suivantes :

1. Scannage du certificat de décès (CMDC). Le certificat « papier » pourra être détruit.
2. Tentative d'appariement de la banque d'image obtenue avec les enregistrements de la bande informatique Insee (reconnaissance optique des caractères). La première clé d'appariement comprend la date du décès, la date de naissance et le sexe. En cas d'échec, une deuxième clé d'appariement comprend le code de commune Insee et le numéro d'acte. Les problèmes d'appariements (10 à 20 % des cas) sont revus par un opérateur qui contrôle la lecture optique et vérifie sur les données de la bande Insee (en cas de problème d'appariement persistant, un retour vers l'Insee est possible).

Le bulletin de décès n° 7, reçu avec le certificat de décès n'est pas saisi ; il sert uniquement aux vérifications en cas de problème d'appariement.

3. Saisie vocale des causes de décès (2,5 à 3 fois plus rapide que la saisie manuelle),

4. Codage des causes de décès, réalisé automatiquement par le logiciel Styx, développé par le CépiDc et spécifique à l'Inserm. La Cim 10 est utilisée depuis 2000 (10 000 libellés au lieu de 5 000 dans la Cim 9 utilisée auparavant). Ce logiciel comporte environ 25 000 entrées (par exemple, 25 entrées pour les équivalents de « mort subite », 100 pour « infarctus », etc.). Dans 30 % des cas, des libellés complexes ou des termes nouveaux nécessitent une revue « manuelle » du codage. En cas de terme non « reconnu », le codage est décidé par 3 personnes et le texte exact ainsi que le code attribué sont gardés en mémoire. Le répertoire des termes est donc évolutif, il est enrichi au fur et à mesure des nouveaux termes rencontrés.

En cas de problème de codage, l'Inserm peut être amené à contacter le médecin signataire du certificat de décès (retrouvé grâce au tampon appliqué sur le certificat ou en retournant à la mairie déclaration du décès).

5. Choix de la cause initiale de décès par un autre logiciel, suivant un algorithme de décision international qui vérifie la logique de la séquence des différentes pathologies listées par le médecin.

Le croisement des données du fichier envoyé par l'Insee et des certificats médicaux reçus par l'Inserm permet en outre un contrôle de l'exhaustivité des déclarations de décès.

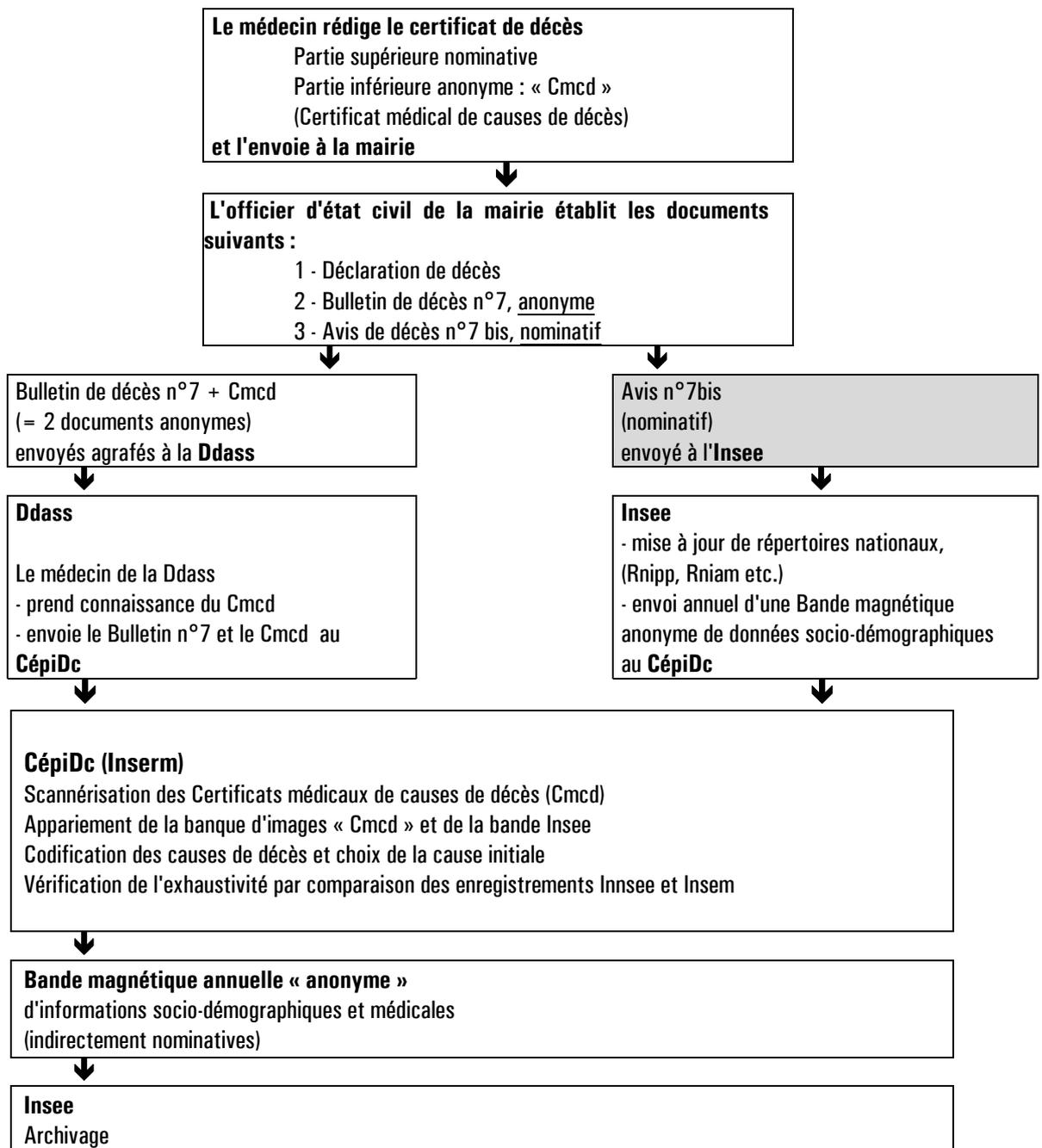
Au total, une **base de données annuelle non directement nominative** est constituée, comportant les données socio-démographiques issues de la base Insee et le codage des causes de décès effectué par l'Inserm. Cette base est disponible 2 à 3 ans après l'année de décès concernée (par exemple, la base des décès survenus en 1999 est finalisée depuis mars 2002).

Cas particuliers :

- Décès dans les Dom-Tom : avant 1999, l'Inserm recevait les certificats de décès mais l'Insee ne recevait pas les notifications de décès.
- Français décédés à l'étranger : l'Insee reçoit les notifications de décès mais l'Inserm ne reçoit pas les causes de décès.
- Étrangers décédés en France : tous les décès survenus en France suivent le même circuit.
- Décès déclarés par un Institut médico-légal : les causes de décès ne sont pas toujours communiquées.

Documentation	Site Internet : http://sc8.vesinet.inserm.fr:1080
Contacts	Réunion avec Eric Jouglu et Gérard Pavillon (CépiDc)

Circuit et traitement du certificat de décès



C.2 Recherche du statut vital des personnes dans le domaine médical

C.2.1 Recherche du statut vital à partir d'informations nominatives

La recherche du statut vital d'une personne peut être effectuée :

1. Auprès des mairies qui sont informées des décès ayant eu lieu dans la commune, ainsi que des décès des personnes habitant la commune ou qui y sont nées. Toute personne qui en fait la demande peut avoir accès au registre des décès et à toutes les informations qu'il comporte.

2. Auprès de l'Inserm

La procédure de demande auprès de l'Inserm est régie par le décret n° 98-37 du 16 janvier 1998 concernant « l'accès aux données relatives au décès des personnes inscrites au répertoire national d'identification des personnes physiques dans le cadre des recherches dans le domaine de la santé ». Ce décret autorise l'utilisation des données figurant au Rnipp en vue d'effectuer « des traitements automatisés de données nominatives ayant pour fin la recherche dans le domaine de la santé... ». Une autorisation délivrée par la Cnil est un préalable obligatoire à une recherche selon cette procédure.

C'est un service de l'Inserm, le Centre de ressources informatiques de l'Ifr69 (Cri-Ifr69) qui reçoit et gère les demandes.

La demande peut porter uniquement sur le statut vital, seul ou associé aux causes de décès. Elle doit être faite à l'aide d'un fichier au format standardisé sur lequel figurent nom (nom de jeune fille pour les femmes), prénom, date de naissance, lieu de naissance et sexe (en cas de petit nombre de cas, la demande peut se faire sur des fiches papier). Le numéro de Sécurité sociale n'est pas communiqué.

Le Cri, après s'être assuré de l'autorisation de la Cnil, transmet le fichier de demande d'identification (ou les fiches manuelles le cas échéant) à l'Insee sans mentionner l'identité du demandeur car l'objet de la recherche est confidentiel.

L'Insee, après appariement (à l'aide d'un logiciel « maison ») de ce fichier avec le Répertoire national d'identification des personnes physiques (RNIPP), renvoie le fichier au Cri dans un délai de moins d'une semaine, après y avoir porté les mentions suivantes (le cas échéant) : date et lieu de décès, numéro d'acte de décès ainsi que des indicateurs de « certification de l'état civil » (qualité d'appariement de l'état civil). La procédure d'identification permet en effet une identification large des « voisins » (personnes ayant des identités proches). Le taux d'identification est de 70 à 99 % selon la qualité du fichier (souvent dépendante de sa taille).

Si la demande porte également sur les causes de décès, le Cri transmet ensuite le fichier reçu de l'Insee au CépiDc après suppression du nom et des prénoms, accompagné du bordereau de transmission reprenant l'identité du demandeur, ceci afin de faciliter les échanges directs entre le demandeur et le CépiDc. L'appariement se fait sur le sexe, la date de naissance, la date et le lieu de décès (département et commune) ainsi que sur le numéro d'acte de décès depuis 1999. La cause initiale, la cause principale et les causes associées du décès sont transmises au chercheur par l'intermédiaire du Cri. En cas de problème, le CépiDc peut prendre directement contact avec le chercheur.

Le délai de réponse est d'environ 1 semaine pour une demande de statut vital simple, de 2 semaines si les causes de décès sont demandées.

Le coût d'une demande est d'environ 30 centimes d'euros par sujet.

Le Cri reçoit une vingtaine de demandes par an, portant sur un total annuel d'environ 150 000 sujets. Une demande peut porter sur un nombre très variable de sujets : d'une vingtaine à plusieurs milliers (jusqu'à 50 000 à ce jour).

Les organismes demandeurs peuvent être des organismes publics ou des équipes de recherche (Inserm, Cnts, Ors, Registres, etc.).

Le détail de la procédure de demande de statut vital au Cri est disponible sur le site du Cri.

Documentation	site Internet : http://ifr69.vjf.inserm.fr/svcd.html
Contacts	Philippe Myquel (Cri-Ifr69)

Au total, les équipes travaillant dans le domaine médical (recherches cliniques ou épidémiologiques, registres etc.) qui souhaitent connaître le statut vital de personnes doivent :

⇒ soit recourir à des méthodes « artisanales » de recherche :

- demande auprès du médecin traitant, consultation des dossiers médicaux,
- demande écrite aux mairies,
- consultation mensuelle auprès des Ddass des bulletins n° 7 de décès et des certificats de décès et tentative de croisement manuel avec des données non directement nominatives.

⇒ soit faire une demande d'identification au Rnipp via le Cri.

Cette démarche nécessite de constituer un dossier de demande à la Cnil spécifique de l'étude concernée et a un coût non négligeable si le nombre de sujets concernés est élevé.

C.2.2 Recherche du statut vital en l'absence d'informations nominatives

Il n'existe pas actuellement en France de système d'information permettant de connaître le statut vital des personnes incluses dans des bases de données hospitalières anonymisées telles que le Pmsi et le Dmi2.

Le Programme de médicalisation des systèmes d'information (PMSI)

Dans le cadre du Pmsi (séjours Mco et Ssr), les décès intra-hospitaliers sont codés et figurent dans les tableaux d'activité des hôpitaux et dans les sorties décès par Ghm. Ils peuvent donc être reliés aux pathologies codées pendant le séjour hospitalier au cours duquel a eu lieu le décès (ou aux précédents le cas échéant, depuis que les séjours hospitaliers peuvent être chaînés).

Par contre, les données hospitalières issues du Pmsi n'étant disponibles que sous forme de Résumés de sortie anonymes (RSA), il est actuellement impossible de les relier à une base de mortalité selon les procédures qui viennent d'être décrites.

Le Dossier médico-économique et épidémiologique de l'immunodéficience humaine (DMI2)

Le Dmi2 est un traitement automatisé d'informations nominatives. Mis en place à partir de 1992, il est implanté dans des centres de référence hospitaliers de l'infection à Vih, les Centres d'information et de soins de l'immunodéficience humaine (CISIH). Il couvre l'ensemble des activités hospitalières : court séjour, hôpital de jour et consultations externes. La direction des Hôpitaux est chargée du traitement médico-économique du Dmi2. L'Inserm (Emi 0214) exploite les données relatives au volet médico-épidémiologique.

La base centralisée par l'Inserm est anonyme et il est donc impossible d'avoir recours au Centre de ressources informatiques de l'Inserm pour connaître le statut vital des patients. Quand les décès sont connus par les services cliniques qui suivent les patients, la date de décès et la cause sont enregistrées au niveau des hôpitaux, sans précision sur le lieu de décès, intra ou extrahospitalier.

Conclusion

En France les modalités d'accès aux données de mortalité individualisées nécessitent donc de disposer d'informations nominatives. Connaître en routine le statut vital d'individus suite à leur prise en charge par le système hospitalier, via les Résumés de sortie anonymes (RSA) est aujourd'hui impossible en France.

D. Systèmes de chaînage à l'étranger

Plusieurs pays disposent de systèmes d'information leur permettant de chaîner une base de mortalité à d'autres bases, administratives et issues du système de santé (résumés de sortie hospitaliers, registres de santé, études cliniques ou épidémiologiques etc.)^[89].

Il peut s'agir :

- soit de systèmes de chaînage organisés à un niveau national ou régional, généralement gérés par des organismes publics,
- soit de chaînages réalisés ponctuellement pour des études ad hoc.

Nous présenterons à titre d'exemple quelques systèmes de chaînage parmi ceux les plus souvent cités dans la littérature ou dont les résultats figurent dans les comparaisons internationales. Cette liste n'est pas exhaustive.

Ces informations ont été obtenues au cours du dernier trimestre 2002 :

- par une recherche bibliographique sur Medline (mots-clés : « death certificates », « medical record linkage », « linkage procedure », « mortality », « outcome assessment », « information systems », « patient identification systems », etc.),
- complétée, lorsque cela a été possible, par des échanges directs avec les responsables de ces systèmes (e-mail ou entretien téléphonique).

D.1 Grande-Bretagne

D.1.1 Écosse

Les premières expériences de chaînage ont débuté en Écosse à la fin des années 1960 pour des études ad hoc essentiellement épidémiologiques (entre 1968 et 1986, une quarantaine de chaînages ont ainsi été effectués). Un système de chaînage permanent a été mis en place à partir de 1989.

Pays (nom du système)	Écosse (« Scottish Record Linkage System »)
Date de début	1989 (1968)
Administration	« ISD » (<i>Information and Statistics Division</i>), Scottish Health Service
Bases de données Chaînables (2 bases permanentes et des chaînages ad hoc)	<p>⇒ 2 bases chaînées en permanence :</p> <p>Base A : enregistrements issus des fichiers suivants :</p> <ul style="list-style-type: none"> - données hospitalières du NHS d'Écosse, à l'exclusion de l'obstétrique et de la psychiatrie, - données hospitalières de psychiatrie, - registre des cancers, - registre général des décès. <p>La base A contient les données hospitalières depuis 1975 et porte sur plus de 5 millions de patients, correspondant à plus de 20 millions de contacts avec le secteur hospitalier de court séjour.</p> <p>Base B : enregistrements issus des fichiers suivants :</p> <ul style="list-style-type: none"> - maternités, - registre des naissances, - services de néonatalogie, - registre général des mort-nés et des décès infantiles, <p>Cette base couvre la période 1981-1999. Un numéro Mère permet de relier entre eux tous les enregistrements d'obstétrique et de néonatalogie se référant à la même femme.</p> <p>⇒ Par ailleurs, des chaïnages ad hoc avec des données externes peuvent être réalisés sur les bases permanentes.</p>
Technique de chaînage	<p>Le chaînage est réalisé à l'aide d'un algorithme de chaînage probabiliste (après chiffrement par Soundex/NYSIIS).</p> <p>Identifiants utilisés pour le chaînage : nom, sexe, date de naissance, lieu de résidence etc. (le numéro d'hospitalisation peut être utilisé en cas de chaînage intra ou inter hospitalier).</p> <p>La performance de l'algorithme de chaînage a été estimée à environ 98 % pour la base A.</p>
Confidentialité	<p>Le chaînage est effectué à partir d'identifiants nominatifs mais une réglementation très stricte définit qui peut accéder aux données et sous quelle forme (sans restriction, sous forme anonymisée ou agrégée). Le plus souvent, les données sont anonymisées (après le chaînage) avant d'être transmises.</p> <p>Chaque demande de chaînage avec des données externes doit faire l'objet d'une demande auprès du « Privacy Advisory Committee » qui évalue l'intérêt de l'étude et donne son avis sur l'opportunité du chaînage et sur les restrictions éventuelles à apporter au moment du rendu des résultats (anonymisés, agrégés etc.).</p>

Applications	<p>A) Des milliers d'analyses ont été réalisées à partir des bases de données chaînées en permanence :</p> <ul style="list-style-type: none"> - études épidémiologiques, - analyse de la mortalité post-hospitalière, des taux de réadmissions, des hospitalisations multiples, de l'évolution des admissions en urgence, - indicateurs de résultats cliniques (à un niveau national ou par « trust » d'hôpitaux). <p>Parmi la quarantaine d'indicateurs de résultats cliniques diffusés, certains concernent la mortalité :</p> <ul style="list-style-type: none"> - taux de mortalité post chirurgie (groupe de 11 interventions chirurgicales) : taux de mortalité à J30, quelle que soit la cause de décès, brut et après standardisation par âge, sexe et niveau socio-économique (« deprivation category ») sur la population de l'ensemble des patients ayant subi l'intervention entre 1992 et 2000 et par type d'intervention. Les résultats sont présentés par « <u>trust</u> » <u>d'hôpitaux</u> (et non par hôpital) par période de 3 ans. - survie à J30 après une admission en urgence pour infarctus du myocarde ou un accident vasculaire cérébral, - survie à J30 et J120 après admission pour fracture de hanche, - survie à 1 an et à 5 ans après cancer du poumon, - taux de survie brut et après standardisation par âge, sexe et niveau socio-économique sur la population Écosse entière. <p>Les résultats, diffusés sur Internet sont encadrés par des précisions méthodologiques et précédés d'un avertissement sur le mode de lecture et d'interprétation. Ils sont présentés sous forme de tableaux et de graphiques de comparaison par rapport à la moyenne du pays.</p> <p>B) Par ailleurs, de nombreux chaînages sont réalisés sur ces bases permanentes avec des données externes (environ 150 sur une période de 5 ans) : essais cliniques, études épidémiologiques, études de surveillance, audits cliniques, etc.</p>
Sites Internet	<p>Information and Statistics Division : http://www.show.scot.nhs.uk/isd Indicateurs hospitaliers : http://www.show.scot.nhs.uk/indicators/outcomes</p>
Références	[67] [68]

Exemples d'études réalisées à l'aide du Scottish Record Linkage System

- Accidents vasculaires cérébraux et mortalité à 6 mois : étude portant sur les données hospitalières de 5 hôpitaux (4 223 admissions, 4 017 dossiers audités, 2 724 Avc et 1 007 décès) ^[141].
- Infarctus et mortalité à J30 : étude réalisée sur une cohorte portant sur 40 371 admissions pour infarctus et 18 452 décès hors hôpital ayant le diagnostic d'infarctus comme 1^{ère} cause de décès ^[14].
- Woscops : essai thérapeutique randomisé (pravastatine contre placebo) sur la prévention de la maladie coronaire portant sur 80 184 hommes de 45 à 64 ans.
- Insuffisance cardiaque et mortalité intra-hospitalière à J30 et à 1 an : étude portant sur 29 hôpitaux et 31 452 patients ^[133].
- Chirurgie prostatique et mortalité tardive : Comparaison de 2 techniques chirurgicales ^[50].
- Évolution des décès par pathologie coronarienne (avec ou sans hospitalisation) entre 1986-1995 : cohorte rétrospective sur toute l'Écosse, incluant les cas hors circuit hospitalier ^[15].

Documentation	Recherche sur medline et site Internet Documents adressés par Steve Kendrick (ISD)
Contacts	Échanges par mail avec plusieurs personnes de l'ISD

D.1.2 Angleterre

Le Nhs diffuse des indicateurs de performance hospitaliers pour l'ensemble du territoire de l'Angleterre.

Pays	Angleterre
Date de début	Le rapport de février 2002 sur les indicateurs de performance du NHS, le 3 ^{ème} du genre, est le premier à comporter des indicateurs hospitaliers.
Administration	Département de la Santé et « CHI » (<i>Independent Commission for Health Improvement</i>). À partir de 2003, la CHI sera responsable de la publication de ces indicateurs.
Bases de données Chainables	- données hospitalières - données de mortalité (qui dépendent de l' « Office for National Statistics »)
Technique de chaînage	Méthode déterministe en 4 étapes sur les variables suivantes : numéro NHS, Date de naissance, sexe et code postal.
Indicateurs de résultats des soins	Les indicateurs sont regroupés en 6 rubriques (d'après le « NHS Performance Assessment Framework » - PAF). Dans la rubrique « indicateurs de résultats des soins » (health outcomes of NHS care) : - mortalité à J30 après chirurgie (programmée et en urgence), - mortalité à J30 après pontage coronarien (calcul sur données de 3 années), - mortalité à J30 après fracture de hanche, - mortalité à J30 après accident vasculaire cérébral, Les résultats, diffusés sur internet, sont rendus par <u>trust d'hôpitaux</u> : nombre de décès (intra ou post hospitalier) pour 100 000 patients après standardisation sur l'âge et le sexe (avec intervalles de confiance). Évolution dans le temps, comparaison à la moyenne du pays (classement sur une échelle de 1 à 5). Les finalités, les modalités de calcul et de standardisation sont détaillées.
Sites Internet	http://www.doh.gov.uk http://www.doh.gov.uk/nhsperformanceindicators/2002

Documentation	Site Internet
----------------------	---------------

D.1.3 Région d'Oxford

Dès 1963 un système de chaînage a permis la mise en place d'une grande base de données épidémiologique dans la région d'Oxford. Limitée au départ à une partie de la région, elle a été progressivement étendue à la totalité des 8 districts (1987).

Pays (nom du système)	Région d'Oxford (« Oxford Medical Record Linkage Study »)
Date de début	1963
Administration	« UHCE » (<i>Unit of Health-Care Epidemiology</i>), département de santé publique de l'Université d'Oxford (Institute of Health Sciences)
Bases de données Chaînables	<ul style="list-style-type: none">- données hospitalières (hôpital de jour inclus) chaînables entre elles,- registre des naissances,- registre des cancers ,- certificats de décès et causes de décès (disponibles pour les résidents de la zone quel que soit le lieu de décès mais non pour les personnes ayant définitivement déménagé de la zone (~ 2% de la population par an). La base porte sur une population de 2,5 millions de personnes (10 millions d'enregistrements en 1997).
Technique de chaînage	« OX-LINK » : Méthode de chaînage probabiliste avec traitement orthographique « ONCA » (Oxford Name Compression Algorithm, une adaptation de la méthode Soundex/NYSIIS) sur nom, 1 ^{ère} lettre du prénom, date de naissance etc.
Confidentialité	Anonymisation des données chaînées avant analyse
Applications	<ul style="list-style-type: none">- études épidémiologiques,- recherche clinique,- organisation sanitaire.
Site Internet	http://www.uhce.ox.ac.uk/index.html
Références	[41] [42] [43]

Exemples d'études :

- Interaction entre traitements cardio-vasculaires et décès à J30 après chirurgie programmée ou en urgence sous anesthésie générale ^[124].
- Facteurs de risque de décès par cause cardio-vasculaire à J30 après anesthésie et chirurgie en urgence ^[55].
- Mortalité post-chirurgicale précoce (<J30) et tardive (90 à 364 jours) ^[123].
- Facteurs de risque de décès par cause cardio-vasculaire à J30 après chirurgie programmée sous anesthésie générale - 1979 à 1990 ^[56].
- Évolution entre 1963 et 1998 de la mortalité post-chirurgicale (intra-hospitalière, après transfert et hors de l'hôpital) à J30 ^[42].
- Étude internationale sur Avc et décès à 6 mois ^[142].
- Étude sur les associations de certaines pathologies (cancer, schizophrénie, maladies auto-immunes etc.) ^[41].

Documentation	Recherche sur medline et site Internet
----------------------	--

D.2 Canada

D.2.1 Generalized Record Linkage System

Le Generalized Record Linkage System (GRLS) a été développé au Canada au niveau national il y a une quinzaine d'années.

Pays (nom du système)	Canada : Generalized Record Linkage System (GRLS)
Date de début	~ 1987
Administration	« Statistiques Canada » (SC)
Bases de données chaînables	<p>1 - chaînage permanent entre les données hospitalières, 2 - chaînages ad hoc de 2 ou plus des fichiers suivants :</p> <ul style="list-style-type: none"> - fichier mortalité (données collectées sous la responsabilité des autorités des provinces ou des territoires et centralisées par Statistiques Canada qui gère la base à des fins statistiques), - registre des cancers, - registre des naissances, - fichier spécifique de la population étudiée : groupe professionnel, cohorte de personnes atteintes d'une pathologie par exemple . <p>Le chaînage entre données hospitalières et données de mortalité ne peut être réalisé que par certaines provinces (le numéro utilisé par les hôpitaux n'est pas toujours disponible dans les bases de décès).</p>
Technique de chaînage	<p>Logiciel développé par Statistique Canada : Generalized Record Linkage Software (GRLS), chaînage probabiliste intégrant une méthode de chiffrement (dans une étude de validation des méthodes de chaînage sur les données de naissances et de décès, le taux de chaînage des décès néonataux était > 99 %).</p> <p>Les données hospitalières, non nominatives sont référencées par le « health insurance number » attribué dans chaque province ou territoire (ce qui interdit le suivi en cas de déménagement vers une autre province) ⇒ certains chaînages ne peuvent être réalisés qu'à un niveau régional (la correspondance entre nom et numéro est conservée dans chaque province). Les données mortalité et cancer sont reçues par Statistique Canada de façon nominative.</p>
Confidentialité	Le couplage sert à des fins statistiques ou de recherche et est conforme au mandat de Statistique Canada défini dans la « Loi sur la statistique ». La diffusion des produits du chaînage satisfait aux dispositions relatives à la Loi sur la confidentialité de la statistique ainsi qu'aux dispositions applicables de la loi sur la protection des renseignements personnels.

	<p>Les projets de chaînage doivent satisfaire aux exigences d'un processus d'examen et d'évaluation qui comprend la présentation à un comité d'experts interne à Statistique Canada puis la soumission au ministre et si nécessaire au commissaire à la protection de la vie privée, voire à des groupes de défense de l'intérêt public et des membres du public lorsqu'il s'agit de nouveaux types de couplage. Seuls les couplages approuvés peuvent être mis en œuvre. Ceux qui sont entrepris doivent être déclarés dans un rapport annuel au parlement.</p> <p>Pour certaines études, le chaînage ne peut être effectué que pour les personnes qui ont donné leur consentement. Les statistiques produites ne doivent pas permettre d'identifier les individus.</p> <p>La durée de conservation de la base chaînée dépend du contexte de l'étude ; les données sont détruites lorsque le projet prend fin.</p>
Applications	<ul style="list-style-type: none"> - Suivi de la mortalité, des cancers ou des naissances dans le cadre d'études de cohorte ou études cas-témoins, d'essais etc., - Statistiques de survie, - Mise en place, maintenance et contrôle d'exhaustivité des registres (cancer, HIV etc), - Incidence régionale de pathologies, - Contrôle de la qualité des bases de données (recherche de doublons, données manquantes...) - Études de surveillance (nutrition par exemple), - Étude sur la santé professionnelle et environnementale, - Études des facteurs influençant l'utilisation et le coût des soins. <p>Pas d'indicateurs de performance par établissement. Les résultats sont présentés au niveau national ou des provinces et non par hôpital Depuis 2002, mise en place de 14 champs « d'indicateurs comparables sur l'état de santé et sur la performance des systèmes » au niveau national et/ou des provinces (PIRC : Performance Indicators Reporting Committee). Dans ce cadre, diffusion par exemple pour 4 provinces (Nouvelle Ecosse, Nouveau Brunswick, Alberta, Colombie-Britannique) des taux de survie à J180 après un accident vasculaire cérébral et à J365 après un infarctus du myocarde (données annuelles pour la province : taux selon le sexe chez les personnes de 45 ans et plus, standardisés sur l'âge, avec intervalle de confiance).</p>
Sites Internet	<p>Statistiques Canada : http://www.statcan.ca</p> <p>ICIS (Institut Canadien d'Information sur la Santé) : http://.cihi.ca/cihiweb</p>
Références	[32] [95] [130] [131]

Canada : exemples de projets approuvés en 2002 (extraits)

- Étude rétrospective par cohorte des risques de cancer chez les infirmières et infirmiers de la Colombie-Britannique.
- Étude sur les anciens combattants de la guerre du Golfe.
- Étude de la mortalité et de l'incidence du cancer chez les travailleurs de l'industrie de l'aluminium de la Colombie-Britannique.
- Le point sur l'étude de la santé des immigrants au Canada: tendances en matière de mortalité et incidence du cancer chez les immigrants.
- Comité chargé de faire rapport sur des indicateurs de rendement : données sur les patients hospitalisés pour des maladies cardiaques ou cérébro-vasculaires.
- Études et contrôle permanent de la mortalité et de l'incidence du cancer chez les travailleurs de la fonderie d'aluminium Alcan au Québec.
- Bilan de l'étude sur la mortalité et l'incidence du cancer chez les travailleurs des établissements industriels de Dow Chemical Canada.

Exemples d'études publiées

- Santé professionnelle : étude épidémiologique portant sur 16 000 mineurs uranium ^[97].

- Herbicides et risque de décès par cancer de la prostate : cohorte rétrospective portant sur 2 213 478 personnes années et 1148 décès par cancer de la prostate ^[90].
- Cancer du cerveau et mortalité : étude de cohorte portant sur 156 242 fermiers ^[91].

Documentation	Recherche sur medline et site Internet
Contacts	Échanges par mails avec Martha Fair (Statistiques Canada)

D.2.2 Système de Manitoba

Parallèlement au système national développé par « Santé Canada », des provinces comme le Manitoba et la Colombie britannique ^[17] ont développé leur propre système de chaînage.

Nous présenterons celui du Manitoba, souvent cité. Cette province compte plus d'un million d'habitants dont 60 % sont concentrés dans la capitale provinciale de Winnipeg où sont implantés 2 hôpitaux universitaires.

Pays (nom du système)	Canada, province du Manitoba : « Population Health Research Data Repository »
Administration	« MCHP » (<i>Manitoba Center for Health Policy</i>), faculté de médecine de l'université de Manitoba (« Department of Community Health Sciences »)
Bases de données chaînées	<ul style="list-style-type: none"> - Base de données de l'assurance maladie (« Manitoba Health insurance registry ») et bases médicales ambulatoires et issues des établissements de santé - Statistiques de décès - Registre des cancers - Fichiers de recherche, comme par exemple celui de l'étude longitudinale sur le vieillissement (« Manitoba Longitudinal Study on Ageing ») <p>Les données sont disponibles depuis 1970 et couvrent l'ensemble de la population.</p>
Confidentialité	<p>La confidentialité est préservée par :</p> <ul style="list-style-type: none"> - la suppression des noms et des adresses des patients et des médecins avant l'envoi des données au MCHP, - la perturbation des données, - le contrôle de l'accès physique aux données. <p>Les procédures de sécurité sont vérifiées par un audit de sécurité. Seuls les projets dont le sujet de recherche a été jugé crédible et utile et ayant eu l'aval d'instances d'éthique (« Manitoba Health Access and confidentiality committee ») sont acceptés.</p>
Applications	<ul style="list-style-type: none"> - épidémiologie, - organisation des soins, - indicateurs de santé et d'utilisation des services par région (par exemple, taux de pontage coronarien pour 1000 habitants selon les zones géographiques ; mortalité prématurée après pontage coronarien selon le taux d'interventions).
Site Internet	http://www.umanitoba.ca/centres/mchp/index.html
Références	[6] [114] [115] [116]

D.3 Australie occidentale

Un système de chaînage généralisé a été développé en Australie occidentale (dont la superficie représente un tiers de l'Australie et qui compte 1,7 million d'habitants). 73 % de la population habitent la capitale, Perth, où 5 hôpitaux universitaires sont implantés.

Pays (nom du système)	Australie, région Australie Occidentale : « Western Australia Linked Database »
Date de début	1995
Administration	Le système est géré par le « Data Linkage Unit » créé à l'initiative du Département de la Santé (<i>Health Information Center</i>) et de l'Université de l'Australie de l'Ouest (<i>Center for Health Services Research</i>). L'unité qui gère le chaînage est localisée dans les locaux du ministère (<i>Health Information center of the State Department</i>). Elle est financée à 60 % par le département de la Santé, 20 % par des bourses et 20 % par la facturation des demandes de chaînage.
Bases de données chaînables	<p>A – Une Banque de données centrale relie les fichiers suivants :</p> <ul style="list-style-type: none"> - registre des naissances (pour la période 1980-nov 2001 ⇨ ~ 530 000 enregistrements), - registre des décès incluant les causes de décès (1969-jul 2002 ⇨ ~ 305 000 enregistrements), - données hospitalières (1970-jul 2002 ⇨ ~ 12 152 000 enregistrements), - hospitalisations et consultations en psychiatrie (1966-juil 2002⇨ ~ 274 000 patients), - registre des cancers (1981-juil 2002 ⇨ ~ 154 000 patients), - registre des sages-femmes (recensement de tous les accouchements, à l'hôpital et à domicile (1980-juil 2000 ⇨ ~ 1 600 000 enregistrements), <p>B – D'autres bases de données peuvent être impliquées :</p> <ul style="list-style-type: none"> - systèmes d'information dépendant du Commonwealth (Personnes Agées, diabétiques, liste d'attente pour institutions) ou de l'état (urgences, ambulances etc.), - bases de recherche (Etude Monica, accidents de la route, etc.), - « geocoding » des adresses pour les données hospitalières, de mortalité et des sages-femmes. <p>Le système comptait en 2002 environ 14 millions d'enregistrements.</p>
Technique de chaînage	Le chaînage est réalisé par une méthode probabiliste après chiffrement par Soundex/NYSIIS, à partir de variables présentes dans les 6 bases de données : numéro patient unique (seulement dans les hôpitaux publics), nom, prénom, sexe, date de naissance, adresse. En 1996, les pourcentages de chaînages faussement positifs et faussement négatifs étaient estimés à 0,11 %.
Confidentialité	L'unité de chaînage est supervisée par un comité directeur à large représentation, comprenant un membre du conseil des usagers . Les études nécessitant l'utilisation du système de chaînage doivent obtenir l'accord du « Confidentiality of Health Information Committee ». L'accès aux données chaînées est réservé aux chercheurs qui présentent les qualités requises par le comité d'éthique. La confidentialité et la sécurité des données est assurée par des protocoles très stricts. Dans la majorité des cas, la recherche est effectuée sur des données anonymisées (données issues du système de chaînage sans aucune mention de nom, d'adresse ou de date de naissance ni d'identité des médecin ou des structures de santé). Il y a séparation entre la procédure de chaînage (réalisée dans le département de la Santé) et l'utilisation des données chaînées (à l'université).
Applications	<ul style="list-style-type: none"> - Planification du système de santé, - Evaluation du système de santé - Recherches en santé : histoire naturelle et pronostic des maladies, études sur l'utilisation et l'efficacité des soins, sur les résultats cliniques et les effets indésirables etc. - Etudes internationales <p>Création en 1996 du « Safety and Quality of Surgical Care Project » (collaboration entre le Royal collège Australien des chirurgiens, le département de recherches en Santé de l'université de WA et le département de la Santé) grâce à la « Western Australia Linked Database ». Objectifs : décrire l'épidémiologie des pathologies nécessitant un recours à la chirurgie, comparer les résultats aux standards de bonne pratique nationaux et internationaux, évaluer de nouvelles technologies chirurgicales, établir des benchmark standards, recommander et évaluer des changements de pratique. La mortalité est l'un des indicateurs de qualité retenus.</p>
Site Internet	http://www.publichealth.uwa.edu.au/research
Références	[52] [66] [125] [128]

Exemples d'études publiées

- Étude de la prévalence du cancer et de l'utilisation des services hospitaliers ^[11].
- Incidence et prévalence des insuffisances rénales terminales et description des traitements ^[9].

- Comparaison du risque d'interventions répétées après 2 techniques chirurgicales sur l'hyperplasie bénigne de la prostate (étude portant sur 19 598 hommes sur 16 ans) ^[127].
- Hyperplasie bénigne de la prostate : comparaison de la mortalité et du risque de cancer après 2 techniques chirurgicales (cohorte de 19 598 hommes sur 17 ans) ^[53].
- Chirurgie du cancer colo-rectal : incidence, mortalité et résultats après chirurgie (9 673 patients entre 1988 et 1995) ^[126].
- Conséquences des accidents de la route : chaînage entre les données de la police, données hospitalières et décès (142 308 cas entre 1987 et 1996) ^[121].
- Hospitalisation des personnes âgées dans leur dernière année de vie (PA > 65 ans décédées entre 1985 et 1994) : évolution du pourcentage d'hospitalisés, de la durée de séjour etc. ^[10].

Documentation	Recherche sur medline et site Internet. Documents envoyés par Diana Rosman (Data Linkage Unit).
Contacts	Échanges par mails avec Diana Rosman

NB : à côté du système généralisé qui vient d'être décrit, on peut noter une autre expérience de chaînage entre données hospitalières et de mortalité, réalisée dans le cadre d'une étude pilote dans un hôpital de Melbourne ^{[98] [99]}.

D.4 Suède

Pays	Suède
Organismes	Centre d'épidémiologie (EpC), dépendant du ministère de la santé (« Swedish National Board of Health and Welfare ») - Fédération des « Swedish County Councils »
Bases chaînées	<p>Le registre des causes de décès peut être chaîné avec :</p> <ul style="list-style-type: none"> - Les données hospitalières (<i>Hospital Discharge Register</i>) : un registre des infarctus du myocarde a été créé en 1996 grâce au chaînage du Registre des données hospitalières aux données de mortalité (330 000 infarctus enregistrés entre 1987-1996). - Certains « registres nationaux de qualité des soins » (<i>National Health Care Quality Registries</i>). <p>Le Registre National des causes de décès et le Registre National des sorties hospitalières dépendent du Centre d'épidémiologie. Le chaînage est réalisé à l'aide du numéro d'identification personnel du patient (supprimé immédiatement après le chaînage).</p> <p><u>Registres nationaux de qualité des soins</u> : il existe en Suède une quarantaine de ces registres portant sur différentes pathologies ou spécialités (tumeurs, diabète, psychiatrie, neurologie, ophtalmologie, maladies cardiaques et vasculaires etc.) Ils incluent des données diagnostiques, de traitement et de résultat. Ces registres sont créés et gérés par des représentants des professionnels de santé. Ils sont encadrés par la fédération des « Swedish County Councils » et le ministère de la santé (« National Board of Health and Welfare ») qui aident et financent leur création et leur développement. Une de leurs missions est de fournir un outil de mesure de la qualité, basé sur des indicateurs de qualité de procédures et de résultats qu'ils ont définis au préalable. Objectifs de ces registres : surveillance des résultats des traitements et de la prise en charge, évaluation des bénéfices pour les patients, travaux de recherche, repérage de problèmes liés à de nouvelles méthodes de traitement ou de technologies. Ces registres sont gérés dans une optique de formation et de développement et non de contrôle. Les résultats sont diffusés en interne sous forme de statistiques agrégées pour l'ensemble du pays et par service, ce qui permet des comparaisons au cours du temps pour un même service ou entre services (dans ce cas de façon anonyme). Beaucoup de ces registres peuvent être chaînés avec le Registre des causes de décès.</p> <p>(Il existe par ailleurs 12 registres épidémiologiques : cancers, naissances, malformations, données hospitalières, causes de décès, statistiques d'avortements (sans n° d'identification), stérilisations, allaitement, reproduction assistée, AMI, statistiques accidents, EHLASS (partie suédoise d'un système européen de surveillance des accidents domestiques et des loisirs). Ces registres, gérés par le centre d'épidémiologie, couvrent la totalité de la population depuis plusieurs décennies et sont utilisés pour des travaux de recherche, d'évaluation et de planification. Ils incluent le « numéro d'identification unique » des sujets).</p>
Applications	<ul style="list-style-type: none"> - Études épidémiologiques, recherche clinique, évaluation et planification. - Indicateurs de qualité des soins : le ministère de la Santé a choisi 60 indicateurs (<i>Overall quality indicators in health care and medical services</i>) ; certains concernent la mortalité (mortalité à J28 post infarctus du myocarde, mortalité à 3 mois post AVC, mortalité liée à certains cancers). Les résultats sont diffusés au niveau national et des comtés. La mise en place des indicateurs globaux de qualité est faite en collaboration avec les « registres de qualité des soins ».
Site Internet	Ministère de la santé : http://www.sos.se Centre d'épidémiologie : http://www.sos.se/epc/epceng.htm Indicateurs de qualité : http://www.sos.se/mars/engflik.htm
Références	[134]

Exemples d'études :

- Études épidémiologiques réalisées sur Registre de l'infarctus du myocarde (incidence, mortalité etc.)^{[46] [119]}.

- Revascularisation précoce et survie à 1 an chez des patients atteints d'infarctus du myocarde et ayant survécu au moins 14 jours : étude sur une cohorte de 21 912 patients (61 hôpitaux entre 1995 et 1998) ^[132].
- Comparaison de la survie après cancerrectal selon l'hôpital de la zone géographique ^[7].
- Survie à 3 mois après hospitalisation pour 27 diagnostics selon la taille de l'hôpital : chaînage du registre H et décès ^[134].
- Évaluation de la qualité des certificats de décès. Comparaison entre diagnostic principal et causes de décès par chaînage entre le Registre des causes de décès et le Registre Hospitalier^[63].

Documentation	Recherche sur medline et site Internet
Contacts	Échanges par mails avec Annika Edberg

D.5 Finlande

Pays	Finlande
Administration	Département d'épidémiologie et de promotion de la Santé, « National Research and Development Center for Welfare and Health » (ex National Board of Health)
Bases de données chaînées	<p>Dans le cadre du CVDR (Finnish Cardiovascular Diseases Registers) project, chaînage des données concernant les patients hospitalisés (cause principale) ou décédés de maladie coronarienne ou d'AVC à partir des bases de données suivantes :</p> <ul style="list-style-type: none"> - Registre hospitalier (<i>Hospital discharge Register</i>), géré par le « National Research and Development Center for Welfare and Health » (ex National Board of Health) . - Registre des décès (<i>National Death Register</i>), géré par « Statistics Finland ». - Données sur le remboursement des médicaments et les pensions d'invalidité pour maladie cardiovasculaire ou AVC, obtenues de la Sécurité Sociale. <p>(participation au projet WHO MONICA) Chaînage réalisé grâce au numéro d'identification unique personnel attribué à tous les personnes résidant en Finlande.</p>
Confidentialité	Anonymisation après chaînage
Applications	<ul style="list-style-type: none"> - épidémiologie, recherche clinique, - aide pour les hôpitaux et la politique de santé. <p>Les résultats en termes de mortalité après infarctus ou AVC ne sont pas utilisés comme indicateurs de performance hospitalière. Ils sont rendus (par âge et sexe) au niveau national ou par zone géographique (et non au niveau hospitalier).</p>
Site Internet	Http://www.ktl.fi/cvdr/findexsf.htm
Références	[79] [80]
Documentation	Recherche sur medline et site Internet
Contacts	Échanges par mail avec M. Mahonen (département d'épidémiologie)

D.6 États-Unis

Une synthèse de la situation aux États-Unis est apparue difficile, compte tenu du nombre et des spécificités de chacun des États.

Le département de la Santé (*US Department of Health and Human Services*) a un registre recensant les efforts des différents états pour intégrer leurs différentes sources de données de santé. On trouvera une description des systèmes d'information mis en place par chacun des États sur le site du ministère (<http://aspe.os.dhhs.gov/statereg>).

Plusieurs États ont développé des systèmes de chaînage entre différentes bases de données, leur permettant notamment de disposer de chiffres de mortalité post-hospitalière (Californie, Washington, New York, etc.).

En Californie par exemple, dans le cadre du « California Hospital Outcomes Project » mené par l'Office of Statewide Health Planning and Development (OSHPD) un système de chaînage entre les données hospitalières et une base de mortalité a été développé en 1996, afin de disposer des décès jusqu'à 365 jours après la sortie de l'hôpital. Le chaînage est réalisé selon une technique déterministe en plusieurs étapes à partir du numéro de Sécurité sociale, de la date de naissance, du sexe, de la race et du code postal. Dans le cadre de ce projet, des rapports sur la mortalité à J30 après infarctus du myocarde par exemple sont publiés. Ces résultats sont détaillés par hôpital et par comté. Ils sont accompagnés de

précisions sur la méthodologie utilisée et sur les interprétations qui doivent être évitées (en particulier, mise en garde contre une interprétation en terme de « classement » des établissements).

Ils sont diffusés sur internet

(<http://www.oshpd.state.ca.us/HQAD/HIRC/Outcome/HeartAttacks/index.htm>).

Documentation	Recherche sur sites Internet
Sites internet	Http://aspe.os.dhhs.gov/statereg/ (US Department of Health and Human Services) Http://www.nahdo.org/project/index.asp (National Association of Health Data Organizations) Http://www.ehdp.com/vitalnet/datasets.htm (liens vers les différentes bases de données) Http://www.ahrq.gov/news/ulp/ulpinfos.htm (« Integrated State Health Information Systems: Summary of a Workshop for State and Local Health Policymakers »)
Contacts	Echanges par mails avec Rita M. Fravel (Agency for Healthcare Research and Quality - AHRQ)

Rochester Epidemiology Project

Il nous a paru intéressant de décrire le « Rochester Epidemiology Project », un système d'information qui a la particularité de couvrir l'ensemble des prises en charge médicales, hospitalières et ambulatoires pour l'ensemble de la population d'un comté du Minnesota. Rochester et le comté d'Olmsted (assez isolé géographiquement) compte 125 000 habitants. 2 hôpitaux, dont la Mayo Clinic, y sont implantés.

Pays (nom du système)	USA : Rochester et comté d'Olmsted, Minnesota : « Rochester Epidemiology Project »
Date de début	1966
Administration	Département de la recherche en sciences de la santé (<i>Department of Health Sciences Research</i>), Mayo Clinic Rochester. (financé par une bourse attribuée tous les 5 ans par l'Institut national de La santé, le NIH, un équivalent de l'Inserm)
Bases de données chaînées	La base inclut l'ensemble des actes médicaux : <ul style="list-style-type: none"> - données hospitalières, - consultations, - urgences, - maisons de retraites, - résultats de laboratoires , - toutes les correspondances au sujet des patients, - décès. Elle couvre presque la totalité de la population de Rochester et du comté d'Olmsted, soit 125 000 personnes (au total plus de 4,5 millions de dossiers, en incluant les patients venant d'autres comtés). NB : les données sont collectées par la Mayo clinic depuis 1909. Un numéro d'identification (numéro de dossier) est attribué au patient lors de son premier contact avec le système de santé (numéro attribué par un des 2 hôpitaux ; si le patient change d'hôpital, le lien se fait sur le nom, le prénom et la date de naissance).
Confidentialité	Pas d'anonymisation, mais les projets de recherche sont soumis à l'accord d'une instance ad hoc. Depuis 1997, un consentement pour l'utilisation des données à des fins de recherche est demandé aux patients à leur entrée dans le système. Pour la période antérieure à 1997, un courrier a été adressé à tous les inscrits pour leur demander leur accord. Un consentement a été obtenu dans 97% des cas.
Applications	Études cliniques, recherches épidémiologiques, génétiques et économiques
Site Internet	Site Internet : http://www.mayo.edu
Références	[87] [112]
Documentation	Recherche sur medline et site internet
Contacts	Échange téléphonique avec Véronique Roger qui travaille sur le projet

Nous présenterons par ailleurs pour les États-Unis un exemple d'enquête de santé utilisant un système de chaînage au niveau national, le **National Health Interview Survey** (NHIS).

Le NHIS est une enquête de santé annuelle portant sur un échantillon représentatif de la population américaine randomisé de 122 000 personnes par an (42 000 foyers). Elle permet un recueil de données sur la santé très vaste : pathologies aiguës et chroniques, utilisation du système de santé (consultation médicales, séjours hospitaliers), espérance de vie, taux de mortalité et facteurs de risque (impact du statut économique, de la couverture sociale) etc. La base de l'enquête proprement dite est chaînée à une base de mortalité.

Pays (nom du système)	USA : « National Health Interview Survey »
Administration	Centre national de statistiques de santé (<i>National Center for Health Statistics</i>)
Bases de données chaînées	<ul style="list-style-type: none"> - BHD (<i>Basic Health and Demographic questionnaire</i>) : questionnaire portant sur la santé, l'utilisation du système de soins et des données démographiques sur toutes les personnes du foyer - CHD (<i>Current Health Topic</i>) : surveillance sur un domaine particulier appliqué à un sous-échantillon (42 000 personnes) - NDI (<i>National Death Index</i>) qui recense tous les décès aux USA depuis 1979 (causes et date de décès renseignées)
Technique de chaînage	Chaînage probabiliste réalisé à partir de 12 combinaisons différentes entre les variables suivantes : n° de sécurité sociale, date de naissance, prénom, « middle initial », nom (et nom du père pour les femmes)
Applications	<ul style="list-style-type: none"> - surveillance de pathologies aiguës et chroniques, - études sur l'utilisation du système de santé (consultation médicales, séjours hospitaliers), - recherches épidémiologiques : espérance de vie, taux de mortalité et facteurs de risque (impact du statut économique, de la couverture sociale etc.).
Références	[104]

Ces quelques exemples ont permis d'illustrer le fait que plusieurs pays, tant en Europe qu'outre-Atlantique, se sont donné les moyens de disposer de systèmes d'information permettant de relier une base mortalité à d'autres bases de données, notamment hospitalières. Les objectifs de ces systèmes varient selon les pays. La mise à disposition de la mortalité post-hospitalière permet d'éclairer différentes problématiques :

- Études épidémiologiques, recherche clinique,
- Recherche en santé publique, évaluation des pratiques,
- Aide aux politiques de santé,
- Démarches qualité (indicateurs de résultat, benchmarking).

Cette donnée est par ailleurs mobilisée dans le cadre d'études internationales comparant l'efficacité respective des différents systèmes de santé (Ocdé, etc.).

E. Systèmes d'anonymisation et de chaînage : la situation en France

E.1 Anonymisation

E.1.1 Rappel réglementaire

La législation française en matière d'identification du patient dans les systèmes d'information est fondée principalement sur la loi Informatique et Libertés du 6 janvier 1978 par laquelle a été créée la Commission nationale de l'informatique et des libertés (CNIL), à qui doit être notifié tout projet de traitement d'informations à caractère personnel. D'après cette loi « *sont réputées nominatives les informations qui permettent, sous quelque forme que ce soit, directement ou non, l'identification des personnes physiques auxquelles elles s'appliquent, que le traitement soit effectué par une personne physique ou par une personne morale* ».

Une modification de la loi de 1978 (qui n'était pas adaptée aux problèmes posés par la recherche statistique) a été soumise à l'Assemblée nationale et doit être discutée au Sénat au cours du premier semestre 2003.

Par ailleurs, le Conseil européen a élaboré en 1995 des recommandations et des directives relatives à la protection des personnes vis-à-vis des traitements de données à caractère personnel pour « *réduire les divergences entre les législations nationales et lever les obstacles à la libre circulation des données à caractère personnel à l'intérieur de l'Union européenne tout en protégeant la vie privée et la liberté individuelle* ». Dans cette directive, les données de santé sont classées comme données sensibles (contrairement à la loi du 6 janvier 1978) dont le traitement n'est possible que dans des cas d'exception : les objectifs et les destinataires du traitement, ainsi que les personnes autorisées à effectuer ce traitement sont strictement définis.

Dans cette directive européenne, les « *données à caractère personnel* » (qui se substituent à la notion « d'informations nominatives ») sont définies comme « *toute information concernant une personne physique identifiée ou identifiable (personne concernée); est réputée identifiable une personne qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques, propres à son identité physique, physiologique, psychique, économique, culturelle ou sociale...* ».

En résumé, sont considérés comme des informations nominatives soumises à la réglementation :

- des identifiants directs faisant référence à une information à caractère personnel indiscutable, comme un numéro d'identification, le nom et prénom, etc.
- des identifiants indirects permettant d'identifier une personne « par référence à des listes nominatives ou à d'autres fichiers (par exemple, numéro de téléphone, initiales du nom et du prénom ou numéro d'ordre renvoyant à une liste nominative, même si celle-ci est établie sur support papier et conservée par chaque médecin), soit par recoupement d'informations, (par exemple, date de naissance et lieu de naissance, code postal de résidence, pathologie rare ...) » ^[34].

La Cnil considère que « le traitement de données de santé à caractère personnel est légitime dès lors qu'il a pour finalités de permettre au professionnel de santé de mieux assurer le suivi médical des patients, de faciliter leur prise en charge par les organismes d'assurance maladie obligatoire, de participer aux actions de prévention et de veille sanitaire poursuivies par les autorités de santé et de contribuer aux travaux de recherche médicale ». Dans le cadre de ces finalités, les données de santé à caractère personnel peuvent être communiquées sous certaines conditions à des destinataires et tiers habilités en vertu de la loi (délibération n° 97-008 du 4 février 1997 sur le traitement des données de santé à caractère personnel).

Dans sa délibération du 4 février 1997, la Cnil a estimé que l'utilisation de données de suivi individualisées, c'est-à-dire d'informations qui peuvent être « chaînées », nécessite des procédures d'anonymisation reconnues et évaluées.

Rappel des principaux textes officiels

- Loi n° 51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques.
- Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
- Loi n° 94-548 du 1er juillet 1994 : Loi relative au traitement de données nominatives ayant pour fin la recherche dans le domaine de la santé et modifiant la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
- Directive européenne 95/46/CE du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, applicable en droit français depuis le 24 octobre 1998.
- Délibération Cnil n° 97-008 du 4 février 1997 portant adoption d'une recommandation sur le traitement des données de santé à caractère personnel.
- Décret n° 99-199 du 17 mars 1999 définissant les catégories de moyens et de prestations de cryptologie pour lesquelles la procédure de déclaration préalable est substituée à celle d'autorisation.
- Afnor FD S97-560 « Informatique et santé. Anonymation : glossaire, démarche d'analyse, expression des besoins ».

Un document publié par le Groupement pour la modernisation du système d'information hospitalier (GMSIH)^[38] fait une synthèse de l'analyse réglementaire sur l'identification des individus, les interconnexions de fichiers et la diffusion des données ainsi que sur les normes européennes.

E.1.2 Techniques d'anonymisation

Afin de permettre le rapprochement de données concernant le même patient tout en respectant l'obligation d'anonymat, il est nécessaire de recourir à des techniques d'anonymisation.

On distingue plusieurs techniques :

1. Les méthodes statistiques d'anonymat, reposant sur la **perturbation des données** : substitution de médianes ou de moyennes aux valeurs réelles, regroupement en classes, ajout de données virtuelles, etc. Cette technique présente plusieurs inconvénients :
 - perte d'information et biais statistique,
 - contrôle de qualité des données difficile,
 - incompatibilité avec les règles de bonne pratique des essais thérapeutiques.

2. Les techniques cryptographiques

2-1 Cryptage

L'utilisation de cette technique, auparavant restreinte par la loi pour des raisons de Défense nationale, a été libéralisée en 1999 (avant cette date, le cryptage était considéré comme une « arme »).

Le cryptage est une technique d'anonymisation qui permet de protéger les données pendant leur transmission (et/ou leur stockage) :

- le texte en clair est chiffré par l'expéditeur avant d'être transmis à l'aide d'une clé,

- le destinataire « déchiffre » le message à l'aide d'une clé et récupère le texte d'origine en clair, sans modification.

On distingue les algorithmes « à clé secrète » (la même clé est utilisée pour le cryptage et le décryptage) ou « à clé publique » (la clé de chiffrement est publique, la clé de déchiffrement est secrète).

Le cryptage est donc réversible ; il garantit la confidentialité des données et autorise le contrôle de qualité. Il est utilisé dans le domaine médical pour la transmission de données entre médecins partageant le secret professionnel.

2-2 Hachage

Cette méthode, utilisée depuis 1995, consiste en une compression numérique transformant les données d'origine en un message de taille fixe. Le hachage est appliqué à des variables directement ou indirectement nominatives, communes aux fichiers que l'on veut apparier. Le code obtenu étant toujours le même pour un individu donné, un rapprochement de tous les enregistrements le concernant peut être réalisé.

Le hachage est irréversible (il est impossible de revenir aux informations d'origine à partir du code) et garantit l'anonymat. En raison de sa fiabilité, cette technique est actuellement préconisée par la Cnil pour les systèmes d'appariement de bases de données médicales en France.

2-3 Signature numérique

Cette technique utilise à la fois les techniques de cryptage et de hachage. Elle permet au destinataire d'authentifier la personne émettrice du message et de vérifier l'intégrité du message reçu.

E.2 Techniques de chaînage

Le chaînage est la mise en correspondance d'informations concernant le même individu issues de plusieurs sources.

Il est réalisé à l'aide de variables communes à ces sources, nominatives ou anonymes. À défaut d'un numéro d'identification personnel unique totalement fiable, il est nécessaire d'utiliser plusieurs variables telles que nom, prénom, date de naissance etc. Ces variables ne sont pas spécifiques d'un individu et certaines peuvent changer au cours du temps, être mal orthographiées ou être manquantes. Dans le cas qui nous intéresse, il s'agit d'apparier des fichiers à partir de numéros d'anonymat obtenus par hachage. Si le repère de chaînage est constitué d'un seul numéro anonyme, il y a appariement entre 2 enregistrements en cas de concordance exacte des 2 numéros correspondants. Si l'on dispose de plusieurs variables, 2 techniques peuvent être utilisées :

1. Le chaînage **déterministe** est fait sur la concordance exacte d'identifiants dont le type et le nombre sont définis en fonction des études. On peut exiger la concordance de tous les identifiants ou d'un nombre minimum (par exemple au moins 3 sur les 5 choisis) ^[2] ^[68] ^[73]. La limite de cette technique est qu'elle considère que tous les identifiants ont la même qualité informative, ce qui pose problème lorsqu'un enregistrement d'une des bases est « chaînable » avec plusieurs individus de l'autre base.

2. Le chaînage **probabiliste** ^{[35] [61] [67] [68] [105]} permet de pallier au défaut du chaînage déterministe car il tient compte de plusieurs « traits » d'identification (en clair ou anonymisés) et les pondère en fonction de leur « valeur discriminante », c'est-à-dire de la quantité d'information qu'ils apportent (par exemple, la date de naissance est plus informative que le sexe). Ce n'est donc pas seulement le nombre de variables communes mais également le type de variables qui sont pris en compte. Pour savoir si 2 enregistrements doivent être mis en correspondance, on applique un modèle d'analyse statistique, paramétré en fonction des objectifs du chaînage, qui calcule pour une base donnée des coefficients de pondération pour chaque variable. Des seuils inférieur et supérieur sont définis selon les risques que l'on s'accorde (en fonction de l'objectif et de la qualité de la base de données) au-delà desquels le chaînage est soit rejeté soit accepté. Les enregistrements se situant dans la zone intermédiaire peuvent être revus manuellement.

	Nom	Prénom	Date de naissance	Sexe	
Enregistrement 1	Dupont	François	29/01/40	M	
Enregistrement 2	Dupond	François	29/03/40	M	
Poids	+9	+4	-3	+2	= 12

D'après C. Quantin et col. ([105]) : Exemple de calcul du poids d'une paire d'enregistrements en fonction de la concordance observée pour chaque variable

Deux types d'erreurs peuvent survenir au cours du chaînage :

1. **Erreur d'homonymie**, c'est-à-dire chaînage à tort de 2 individus différents,

Elle peut être due :

- à une erreur de saisie des identifiants utilisés pour l'anonymisation,
- à l'existence de « collision » lors du hachage (le hachage des identifiants de 2 individus génère le même code),
- à une réelle homonymie d'individus partageant par exemple les mêmes identifiants utilisés pour le hachage (par exemple, mêmes nom, prénom, sexe et date de naissance).

2. **Erreur de synonymie**, c'est-à-dire absence de chaînage d'enregistrements d'un même individu (liée à une erreur de saisie, à un changement de nom etc.)

La qualité des informations recueillies dans les bases conditionne la qualité du chaînage. Toute procédure d'anonymisation en vue d'un appariement de données devrait être précédée d'un contrôle à la saisie et de recherches des doublons.

E.3 Appariements sécurisés : exemples d'applications en France

Plusieurs procédures d'anonymisation et de chaînage ont été développées en France, principalement :

1988 : algorithme de San Marco, par la faculté de médecine de Marseille,

1995 : Anonymat, par le Dim du Chu de Dijon en 1995,

1996 : Foin, par le Cessi-Cnamts (appliqué en particulier au Pmsi, à l'observatoire du Rmi et tout récemment au Sniir-Am).

Les procédures d'anonymisation adoptées dans Anonymat puis Foin sont celles actuellement conseillées par la Cnil. Nous en décrivons les principes communs :

Ces 2 applications utilisent l'algorithme public de hachage Standard Hash Algorithm (SHA), recommandé par le National Institut for Standards and technologies (NIST – USA) comme étant l'algorithme du domaine public le plus sûr vis-à-vis des tentatives de déchiffrement.

La fonction d'anonymisation ainsi obtenue présente les qualités suivantes ^[21] :

- non réversible : il est impossible de remonter, à partir du numéro anonyme obtenu, aux informations nominatives de départ,
- faible taux de collision (probabilité d'obtenir le même numéro d'anonymat à partir de données nominatives différentes),
- bon effet avalanche (pour des données nominatives approchantes, numéros anonymes très différents),
- paramétrable : le numéro anonyme est obtenu à chacune des applications à partir d'une clé secrète paramétrée dans cette fonction.

L'algorithme de hachage étant public, il existe un risque de création de table de concordance identifiants-numéros anonymes. Ce risque « d'attaque par dictionnaire » est prévenu par l'utilisation de 2 clés informatiques (spécifiques du projet) : la première est utilisée au moment du hachage des identités par les centres sources, la seconde au moment de la réception des fichiers par le centre de traitement coordinateur (voir page suivante).

Les conditions de création et de gestion des clés sont très strictes. Par exemple, pour la procédure Foin :

- Utilisation de la technique de « fragmentation – redondance – dissémination » (basée sur le schéma à seuil de Shamir) qui permet de sécuriser la clé secrète : la clé est fragmentée en plusieurs images qui prises individuellement ne contiennent aucune information significative. À partir d'un certain nombre d'images (seuil de reconstruction), on peut recalculer le secret de départ. Les autorités ayant choisi le secret ne sont pas destinataires des images de ce secret.
- Séparation des pouvoirs : la clé secrète est choisie par au moins 2 intervenants pour qu'une seule personne ne détienne pas l'intégralité du secret.
- Stockage sécurisé et disponibilité : les moitiés de secret doivent être stockées sur des supports sûrs en plusieurs exemplaires et dans divers endroits,
- Auditabilité et intégrité : la génération du secret doit être faite par une personne de confiance qui ne gardera aucune trace de ces opérations et sera le garant de l'intégrité de l'opération.

Les personnes gérant les clés sont nommément responsables. Les clés sont spécifiques à chaque projet. Des sécurités d'accès physique (mot de passe, locaux sécurisés etc.) sont également définies.

Par ailleurs, afin de diminuer les erreurs de saisie ou d'orthographe lorsque des variables alphabétiques sont utilisées comme identifiants, un traitement orthographique est intégré au logiciel (suppression des caractères spéciaux, transformation des « y » en « i » etc.).

La procédure d'anonymisation se déroule en 2 étapes :

1. **Chaque centre de recueil** des données (centre source) applique la procédure d'anonymat avec une clé de hachage 1, commune à tous les centres sources, ce qui garantit l'anonymat des données exportées. Cette 1^{ère} anonymisation transforme les identifiants en codes anonymes de longueur fixe. Les identifiants peuvent être alphabétiques ou numériques : nom, prénom, nom de jeune fille, date de naissance, numéro de Sécurité sociale etc. Pour respecter les caractéristiques de faible taux de collision et de bon effet avalanche, le numéro anonyme doit comporter un minimum de caractères alphanumériques (dans le cadre du Pmsi, Foin calcule un anonymat de 17 caractères à partir de données de longueur quelconque). Les données nominatives utilisées pour générer le numéro anonyme doivent toujours être identiques pour un même projet afin de permettre le chaînage des différents fichiers.

Identifiant ⇒ hachage + clé 1, commune aux centres sources ⇒ **code anonyme 1**

2. **Le centre de traitement** chargé du chaînage, destinataire des différents fichiers anonymisés, applique une seconde procédure d'anonymat (avant toute tentative d'appariement) avec une clé de hachage 2, connue uniquement du centre de traitement, ce qui garantit la sécurité des informations chaînées vis-à-vis des centres sources.

code anonyme 1 ⇒ hachage + clé 2, connue uniquement du centre de traitement ⇒ **code anonyme 2**

Le numéro anonyme étant toujours le même pour une identité donnée (à condition d'utiliser les mêmes clés), les enregistrements issus des différentes bases de données peuvent être appariés.

L'appariement entre les enregistrements sources est réalisé à l'aide du « repère de chaînage » constitué par le code anonyme 2. Les techniques de chaînage diffèrent selon les projets.

Nous présenterons les principaux dispositifs d'anonymisation et de chaînage développés en France.

E.3.1 Dim du Chu de Dijon

La procédure Anonymat mise au point par le Dim du Chu de Dijon (équipe du Pr. Catherine Quantin) en 1995 a été déclarée auprès de la Cnil et du Service central de la sécurité des systèmes d'information (SCSSI) en mars 1996.

La procédure d'anonymisation (ANONYMAT), qui suit le schéma qui vient d'être décrit, génère un code anonyme par identifiant. Le chaînage est réalisé selon la technique probabiliste « AUTOMATCH » proposée par Jaro ^[61], avec pondération selon la valeur discriminante de chaque identifiant. L'anonymisation ne gêne pas le chaînage et le nombre d'enregistrements chaînés est le même que les données à chaîner soient en clair ou anonymisées. Le taux de collision est très faible ($\sim 10^{-48}$). Une évaluation de la méthode, réalisée sur le chaînage entre un registre du cancer du colon (2 847 patients) et des données hospitalières (334 848 hospitalisations pour cette cause), a conclu à une sensibilité de 97 % et une spécificité de 93 %.

Un traitement orthographique (dérivé de l'algorithme de San Marco ^[136]) est appliqué avant l'anonymisation. L'intérêt de ce traitement (gain en sensibilité mais perte de spécificité) est évalué dans un article à paraître ^[70].

Exemples d'applications d'Anonymat :

- Étude de l'inter file active des personnes atteintes de cancer et suivies dans 3 structures hospitalières en Rhône-Alpes : le chaînage après anonymisation des enregistrements des 3 bases issues du Pmsi 1996 a permis de repérer les patients communs aux établissements et d'estimer le taux de doublons dans chacune des bases administratives.
- Développement d'un réseau périnatal en Bourgogne (pour le recueil régional d'indicateurs en périnatalité) : chaînage des données issues des 20 établissements (dont 4 privés) prenant en charge les femmes enceintes et les nouveaux nés. L'utilisation d'un Rum élargi (âge gestationnel, facteurs de risque psychosociaux) a nécessité la modification des logiciels de traitement des données Pmsi.
- Réseau Hépatite C en Bourgogne.
- Réseau Hépatite C et tentatives de suicide en Franche-Comté.
- Collaboration avec une équipe de génétique du Cnrs de Grenoble (réflexion sur des identifiants « emboîtés » reliant le sujet à ses ascendants et ses descendants).
- Collaboration avec le Luxembourg et la Suisse. L'office fédéral de la statistique Suisse a adopté un processus d'anonymat et de chaînage similaire à celui de Dijon pour apparier les séjours hospitaliers au niveau fédéral ^[8].

Documentation	Recherche sur medline Séminaire de la Société française de statistique « Pratique des appariements sécurisés » (janvier 2003)
Contacts	Catherine Quantin
Publications	[30] [69] [105] [106] [107] [108] [109]
Textes officiels	Déclarations à la CNIL et au Service central de la sécurité des systèmes d'information (SCSSI) – mars 1996

E.3.2 Cessi / Cnamts

Le Centre d'études des sécurités du système d'Information de la Caisse nationale de l'assurance maladie des travailleurs salariés (CESSI/CNAMTS) a développé en 1996 la procédure Fonction d'occultation d'identifiants nominatifs (FOIN) utilisée initialement dans le cadre du PMSI puis adaptée à de nombreuses applications (essentiellement dans le domaine médical mais aussi social) parmi lesquelles le Sniir-Am et l'observatoire du Rmi.

Pmsi

La fonction d'anonymisation Foin a été développée en 1996 par le Cessi/Cnamts, à la demande de la Direction des hôpitaux pour sa mise en place dans le Pmsi établissements privés. Cette procédure a été approuvée et validée par la Cnil et le Service central de la sécurité des systèmes d'information (SCSSI) en 1996. Elle permet, pour reprendre des extraits du texte de la Cnil « effectivement de garantir l'anonymisation des données transmises. En conséquence, ces données ne présentent pas, au sens de l'article 4 de la loi du 6 janvier 1978, un caractère direct ou indirect nominatif ».

Cette fonction d'anonymisation vient d'être étendue au secteur public depuis 2001, permettant le chaînage des différents séjours hospitaliers Mco et Ssr d'un même patient (dans le temps et/ou dans différents établissements) tout en respectant la confidentialité des données.

Cette fonction, basée également sur l'algorithme Sha, permet de rendre anonyme n'importe quel type de données :

- informations directement nominatives (nom, prénom, etc.),
- informations indirectement nominatives : Nir (numéro d'inscription au répertoire national d'identification des personnes physiques), numéro de bénéficiaire (un bénéficiaire étant rattaché à un assuré social), numéro de professionnel de santé, numéro de rente accident du travail, etc.

Les identifiants utilisés pour la procédure d'anonymisation sont les suivants :

- Numéro de Sécurité sociale de l'ouvrant-droit : l'hôpital enregistre le numéro de l'assuré, c'est-à-dire de l'ouvrant-droit et parfois le rang du bénéficiaire (ce dernier élément ne serait pas recueilli de façon systématique dans tous les hôpitaux).
- Sexe du bénéficiaire.
- Date de naissance du bénéficiaire.

Si le numéro de sécurité sociale n'est pas disponible (en cas de non facturation par exemple) Foin est prévu pour fonctionner avec d'autres identifiants : nom, prénom, sexe et date de naissance. Un traitement orthographique (identique à celui utilisé dans l'algorithme de San Marco ^[136]) est intégré à Foin.

Le hachage des différentes variables identifiantes aboutit à **un seul numéro anonyme** de 17 caractères et le **chaînage** est effectué sur la stricte concordance de ce numéro (pas de méthode probabiliste comme dans la procédure du Dim de Dijon).

Ce dispositif permet d'assurer le chaînage des séjours d'un même patient pour les admissions postérieures au 31 décembre 2000.

Ce dispositif est désormais implanté dans tous les établissements hospitaliers (cf schéma ci-contre extrait de la circulaire Dhos-Pmsi n° 106 du 22 février 2001). Il permet la stricte dissociation du circuit administratif (qui utilise le numéro de sécurité sociale) et du circuit de l'information médicale (pas de numéro de Sécurité sociale) : le chaînage est possible sans que les résumés de séjour soient indexés par le numéro de Sécurité sociale qui constitue un identifiant majeur.

Conçu pour fonctionner avec n'importe quel identifiant, il est pour le moment surtout utilisé avec des variables numériques ce qui évite les problèmes d'erreurs orthographiques et d'homonymies vraies. Mais l'utilisation du numéro de Sécurité sociale de l'ouvrant-droit pose 2 problèmes :

- ce numéro peut changer au cours du temps pour un même sujet (bénéficiaire changeant d'ouvrant doit ou devenant lui-même assuré etc.). En cas de changement, le numéro d'anonymat change et le chaînage avec les enregistrements antérieurs n'est pas possible,
- il ne permet pas de distinguer les jumeaux de même sexe.

En outre, le numéro d'anonymat généré dans le Pmsi ne peut pas servir de repère de chaînage si l'on souhaite apparier la base hospitalière avec une base ne disposant pas du numéro de Sécurité sociale de l'ouvrant-droit du patient.

À notre connaissance, il existe encore peu d'expériences d'anonymisation par la fonction Foin sur des identifiants alphabétiques telles que le nom et le prénom (sans utilisation du numéro de sécurité sociale de l'ouvrant-droit). La qualité de l'appariement dans ces conditions n'est pas connue. Sachant que le chaînage est fait sur un seul numéro d'anonymat (résultat de la compression des différents identifiants utilisés) sans application d'un modèle probabiliste, on peut craindre qu'un pourcentage non négligeable d'appariements soient perdus, compte tenu de la non standardisation de la saisie des identifiants (saisie du nom de jeune fille ou du nom marital ou du « nom usuel », modes de saisie des prénoms composés variables - avec ou sans tiret, complet ou initiales) et des risques d'erreur de saisie et d'orthographe.

Documentation	Recherche sur medline et site Internet du Pmsi Documents adressés par la Cnaamts
Contacts	Échanges avec la Cnam-Ts/DSI/Cessi Réunion avec Mr Lazaro Pejsachowicz (Cnamts) et Marie Laure Buffat (Cnamts/Dsi/Cessi)
Références	[140]
Textes officiels	Accord du Scessi du 3 juin 1996 sur le Fonction d'occultation d'Information nominative de la Cnamts Avis favorable de la Cnil du 18 juillet 1996 sur le Fonction d'occultation d'information nominative de la Cnamts Délibération Cnil n° 96 – 061 du 9 juillet 1996 sur l'utilisation du logiciel Foin dans les établissements privés Accord du Scessi du 8 janvier 1997 portant sur la gestion des clés de la Fonction d'occultation d'information nominative de la Cnamts Circulaire Dhos-Pmsi n° 106 du 22 février 2001 relative au chaînage des séjours en établissements hospitaliers. Dossier Cnil du 20 juin 2000 sur l'utilisation du logiciel FOIN dans les établissements publics

Sniir-Am

La seconde application de grande envergure de Foin dans le domaine médical est le Système national d'information inter régimes de l'Assurance maladie (SNIIR-AM). Ce système a été créé par la loi du Code de la Sécurité sociale du 23/12/98 (signature de l'arrêté ministériel le 11 avril 2002).

Les objectifs du Sniir-Am sont de :

- Connaître les « dépenses de l'ensemble des régimes de l'assurance maladie par circonscription géographique, par nature des dépenses, par catégorie de professionnels responsables de ces dépenses et par professionnel de santé ou établissement ».
- Faire un retour d'informations auprès des professionnels de santé sur leur activité, leurs recettes et le cas échéant leurs prescriptions.

Il s'agit de passer de données statistiques agrégées à des données individuelles afin de connaître les trajectoires de soins des patients et d'analyser le comportement des assurés et des professionnels. Auparavant, les données étaient dispersées entre les différentes caisses (16 régimes) et il était impossible de disposer de données exhaustives permettant d'établir des statistiques fiables et complètes. Le Sniir-Am doit permettre de couvrir l'exhaustivité des

dépenses des régimes, d'assurer une meilleure qualité des sorties statistiques et de partager les données avec les acteurs.

Cette base est gérée par le centre national de traitement informatique de la Cnamts.

« Le Sniir-Am est établi à partir des données d'activité et de dépenses détaillées individualisées et chaînées, y compris les données de codage et celles du Pmsi, couvrant le champ des soins de ville, des soins hospitaliers sanitaires et sociaux, publics et privés ».

Il s'agit du chaînage de données hospitalières et ambulatoires :

- exhaustives,
- individualisées par bénéficiaire anonyme,
- médicalisées : données du Pmsi, codes Ald (seul accès pour le moment aux diagnostics en ambulatoire), codage des actes (biologie), médicaments, etc.

L'organigramme du dispositif (document Cnamts) est joint ci-contre.

Les organismes fournisseurs de données sont la Cnamts, la Msa, la Canam, les Régimes spéciaux et la Dhos.

Les données incluses dans le Sniir-Am sont les suivantes :

- organismes de prise en charge,
- décomptes,
- bénéficiaires (n° anonyme, sexe, mois et année de naissance, département et commune),
- prestations servies (nature de l'acte, date des soins et de remboursement, mode de prise en charge, informations médico-administratives),
- professionnels de santé (n° d'identification, spécialité, nature de l'exercice, département et commune),
- établissements de santé (n° Finess, catégorie, activité, Rsa, Informations de séjours etc.),
- données comptables.

La loi précise que *« les données reçues et traitées par le Sniir-Am préservent l'anonymat des personnes ayant bénéficié des prestations de soins »*. Par contre, les informations concernant les professionnels de santé sont nominatives. L'anonymisation des données d'identification du patient est assurée par la fonction Foin qui génère des numéros « anonymes » permanents dans le temps et l'espace dès lors qu'elle est appliquée sur des données identiques.

Le même principe déjà décrit pour le Pmsi est appliqué ici : une double anonymisation interdit la constitution de tables de correspondance. L'algorithme est paramétré par une clef secrète associée à chaque étape afin de garantir le cloisonnement des informations.

Le 1^{er} niveau d'anonymisation est réalisé par chaque organisme fournisseur avant la transmission des données vers le Sniir-Am, à partir des mêmes identifiants que pour le Pmsi (« identifiant bénéficiaire constitué ») :

- numéro de sécurité sociale de l'ouvrant-droit,
- date de naissance du bénéficiaire,
- sexe du bénéficiaire.

NB : Les problèmes liés à l'utilisation du numéro de Sécurité sociale de l'ouvrant-droit ont été évoqués plus haut. Le nom et le prénom n'ont pas été choisis car ils ne figurent pas dans toutes les bases sources concernées.

Une seconde anonymisation est appliquée aux numéros d'anonymat obtenus à l'étape 1 avant l'intégration des données à l'entrepôt national.

Ce second numéro d'anonymat sert de repère de chaînage entre les différentes bases.

Les données d'identification, la mise en œuvre des niveaux d'anonymisation et les clés secrètes utilisées à chacun de ces niveaux doivent être strictement identiques dans le Sniir-Am et le Pmsi.

Sécurité

Des normes d'alimentation et de contrôles de qualité communs à tous les fournisseurs de données ont été définies.

La fonction d'anonymisation Foin est intégrée au produit Camel qui assure la sécurisation des flux de données entre les organismes fournisseurs et le Sniir-Am par un dispositif de signature électronique, de techniques de chiffrement et la conservation de l'historique des échanges.

La base est centralisée dans un entrepôt national unique.

La Cnamts assure la sécurité des données au sein de l'entrepôt lui-même c'est-à-dire entre l'entrée et la sortie des données. Par contre, en amont, les fournisseurs sont responsables de la qualité des données qu'ils transmettent, en aval, les utilisateurs sont responsables des données qui leur sont transmises.

L'authentification des différents partenaires est assurée par un dispositif de signature électronique.

Les fichiers de sauvegarde sont chiffrés.

Accès aux données

Le Sniir-Am doit être accessible à l'ensemble des caisses des différents régimes de base, aux Unions régionales des caisses d'assurance maladie (URCAM), aux Agences régionales d'hospitalisation (ARH), aux Unions régionales de médecins libéraux (URML), au ministère de l'Emploi et de la solidarité et au ministère de l'Agriculture, aux prestataires de soins pour les données concernant leur activité ainsi qu'à certains instituts de recherche.

L'accès aux données est réglementée :

- Une partition des données est possible : national ou régional, médicales ou non, individuelles ou agrégées, professionnels de santé en clair ou anonymes etc.
- 13 profils d'utilisateurs ont été définis qui conditionnent le type et le périmètre de données qui leur est accessible.

- Certains croisements de données sont interdits, par exemple : [code commune+ code pathologie] ou [date de naissance + date des soins] (seuls les médecins conseils y ont accès).
- Il est impossible de procéder à une requête concernant moins de 10 bénéficiaires.
- « L'intégralité des flux d'alimentation du Sniir-Am est tracée et auditable » (CNIL)
- « L'intégralité des accès aux données du Sniir-Am est mémorisée et auditable » (CNIL).

Différents niveaux d'accès aux données sont prévus :

- Médecins conseils des services médicaux des caisses, personnels placés sous leurs responsabilité et agents administratifs des caisses et des Urcam (ceux-ci nommément désignés par les responsables des organismes (directeur ou agent comptable) : accès à l'ensemble des données individuelles mais anonymisées des bénéficiaires et aux données en clair sur les professionnels de santé.
- Urm, Arh, Ddass, Ministères de l'Emploi et de la solidarité et de l'Agriculture : accès aux données sous forme statistique agrégée, sans identification des professionnels ni données individuelles relatives aux bénéficiaires.
- Professionnels de santé : accès aux données relatives à leur activité, leurs recettes ou leurs prescriptions et aux données individuelles anonymisées de leurs patients.
- Autres organismes demandeurs : s'ils ne sont pas listés nommément dans les textes, ils doivent obtenir une autorisation de la Cnil pour rapatrier des données.

Les utilisateurs ne pourront accéder au Sniir-Am qu'après avoir été identifiés et authentifiés par une carte de sécurité ou un mot de passe et que s'ils figurent sur un annuaire géré par la Cnamts.

« Les données relatives aux bénéficiaires de soins seront conservées pendant 2 ans au-delà de l'année en cours et les données concernant les professionnels de santé 10 ans » (rapport Cnil 2001).

État d'avancement du projet : des tests d'appariement étaient en cours au premier semestre 2003.

Documentation	Cnamts/Dsi/Cessi sept 2001. Camel. Anonymisation des données du Sniir-Am basée sur la Fonction d'occultation d'information nominative. Séminaire de la Société française de statistique « Pratique des appariements sécurisés ». 20-21 janvier 2003 22 ^e rapport d'activité 2001 de la Cnil
Contacts	Yvon Merlière (Cnamts)
Textes officiels	Loi du 23/12/98 – Art. L 161-28-1 – du Code de la Sécurité sociale créant le Sniir-Am Avis favorable de la Cnil du 18/10/2001 Avis favorable du Conseil pour la transparence des statistiques d'assurance maladie (COTSAM) du 04/04/2002 Signature de l'arrêté ministériel du 11/04/2002

Observatoire du dispositif du Rmi à Paris

L'« Observatoire des entrées et sorties du dispositif du Rmi à Paris » est un des rares exemples de dispositif d'appariement réalisé en dehors du champ médical. Mis en place à la demande de l'état et de la Ville de Paris par le département d'évaluation des politiques sociales du Centre de recherche pour l'étude et l'observation des conditions de vie (CREDOC), il est géré au niveau du département.

Son objectif est de mieux connaître :

- Le profil des allocataires et leurs parcours antérieurs,
- Le passage dans le dispositif (attentes, motif d'entrée, actions dont ils ont bénéficié...),
- L'effet des actions sur les sorties.

Le dispositif permet de mettre en relation les données issues de 9 organismes :

- Caf de Paris,
- Fichier national de contrôle de la Cnaf (qui permet de vérifier que l'allocataire ne bénéficie pas du Rmi dans un autre département),
- Fichier historique de l'Anpe – national (chômage, emplois recherchés, niveau de diplômes),
- Déclarations uniques d'embauche de l'Urssaf – Île-de-France (emplois dans le secteur privé),
- 4 fichiers du Cnasea – Île-de-France (contrats aidés, stages de formation),
- Cellule centrale de coordination du département de Paris (contrats d'insertion, difficultés rencontrées etc.).

Cet observatoire permet donc de mettre en relation :

1. Des fichiers de gestion administrative des allocations (profil et situation familiale des allocataires, date d'entrée, montant versé, autres ressources etc.),
2. Des informations relatives aux mesures d'insertion, issues en particulier des commissions locales d'insertion (informations sur les allocataires bénéficiant d'un contrat),
3. Des informations relatives aux situations d'emploi ou de recherche d'emploi, via les données Anpe, Urssaf et Cnasea.

Le délai et le champ spatial des données varie selon l'organisme (seul l'Anpe dispose d'une base historique remontant à 1993).

Descriptif du dispositif

La Caf :

1. Extrait l'échantillon de l'étude à partir de ses fichiers de gestion (entrants et sortants du dispositif et échantillon d'allocataires « témoins » ayant ≥ 1 an d'ancienneté).
2. Procède à son anonymisation, à l'aide d'une adaptation du logiciel Foin fournie par le Cessi/Cnamts. En raison des risques d'homonymies, il a été décidé d'anonymiser des identifiants numériques. L'un des organismes ne disposant pas du Nir mais uniquement du matricule Caf, 2 séries de numéros anonymes (« étiquettes ») ont été créées l'une à partir du Nir, l'autre à partir du matricule Caf (la Caf dispose des 2 identifiants).

NB : les numéros d'anonymat sont valides pour une vague d'enquête seulement

3. Transmet la liste référencée par ces 2 séries d'étiquettes au Credoc.

Le Credoc distribue cette liste aux différents organismes de l'observatoire, en choisissant l'étiquette anonyme adéquate (selon l'identifiant dont ils disposent, Nir ou matricule Caf).

Chaque organisme procède à l'anonymisation de son propre fichier (à l'aide de la même clé que la Caf) et l'apparie avec la liste reçue du Credoc. Le fichier de données ainsi sélectionné est ensuite transmis au Credoc.

Le Credoc apparie les fichiers reçus des différents organismes et procède à l'analyse (les numéros d'anonymat sont supprimés après appariement).

La première vague du projet a débuté en septembre 2001. Elle concernait la période novembre 2000 – mai 2001 et portait sur 37 760 dossiers (rapport publié en décembre 2002).

Documentation	Séminaire de la Société Française de Statistique « Pratique des appariements sécurisés ». 20-21 janvier 2003 (Marie-Odile Simon, responsable de l'observatoire Rmi au Credoc)
Contacts	Marie Odile Simon (Credoc)
Référence	« Observatoire des entrées et sorties du Rmi à Paris, compte rendu de la première vague » Isa Aldeghi, avec la collaboration de Marie-Odile Simon – Collection des rapports n° 226, décembre 2002.
Textes officiels	Avis de la Cnil mai 2000

E.3.3 Autres dispositifs

La procédure utilisée par le Dmi2

Le recueil de l'information dans le cadre du Dossier Médico-économique et épidémiologique de l'immunodéficience humaine (DMI2) est assujéti au consentement écrit du patient. La transmission des données se fait à l'aide d'un numéro d'anonymat créé à partir du nom de naissance, du prénom, du jour et du mois de la date de naissance grâce à « **l'algorithme de San Marco** »^[136]. Ce numéro est constitué lors de l'extraction des données hors du service médical. Il garantit l'anonymat et l'unicité du code : le code est toujours le même pour un même sujet, quel que soit le lieu ou la date du codage et permet le chaînage de toutes les fiches de suivi d'un même patient. La formule est non réversible. Des risques de « collision » existent à une probabilité très faible.

L'Inserm, lorsqu'elle reçoit des fichiers des Csih lance une procédure de recherche des doublons. S'il s'agit de faux doublons (par exemple, un même numéro attribué à un homme et à une femme), les enregistrements sont supprimés.

En cas de nécessité de retour au dossier, l'Inserm fournit au Csih un certain nombre de renseignements dont le sexe, la date de naissance, le groupe de transmission du patient et éventuellement la date de Sida, la date de décès si elles existent, et les critères de l'étude pour laquelle la recherche est nécessaire. À partir de la sélection opérée sur ces critères, le Csih recalcule les numéros d'anonymat à partir de leurs données nominatives, de façon temporaire, à l'écran (non imprimable). Cette procédure a été acceptée par la Cnil.

Fin 2001, la base comportait près de 2 millions de fiches de suivi (1 820 000) concernant 89 950 sujets de plus de 15 ans (1 fiche de suivi = 1 recours à l'hôpital et correspond à toutes les hospitalisations classiques ou de jour, ainsi qu'aux consultations pour lesquelles un événement clinique, biologique ou thérapeutique a eu lieu).

Documentation	Mission Sida. Un système d'information centré sur les patients : le Dmi2 - novembre 1992 – (Document Direction des hôpitaux)
Contacts	Murielle Mary-Krause (Inserm EMI 0214)
Textes officiels	Accord Cnil 10 septembre 1991 Arrêté du 27 novembre 1991 autorisant le traitement informatisé des dossiers médico-économiques et épidémiologiques de l'immuno-déficience humaine dans les centres d'information et de soins de l'immunodéficience humaine et autres établissements hospitaliers

Enquête Santé et protection sociale (ESPS)

Cette enquête est réalisée par le Credes depuis 1988 à la demande et en collaboration avec la Cnamts. Elle a pour objectif de « décrire et d'analyser les relations entre consommation de soins, état de santé, caractéristiques socio-économiques et couverture sociale ». Actuellement, elle est réalisée tous les 2 ans à partir de l'Échantillon permanent d'assurés sociaux (EPAS) de la Cnamts et plus récemment, à partir des échantillons similaires de la Canam (1994) et de la Msa (1996). Elle permet de recueillir des informations sur les caractéristiques individuelles des assurés contactés et de tous les membres de leur foyer : données démographiques et socio-économiques, morbidité, couverture complémentaire, consommation médicale, hospitalisation.

Le service informatique de la Cnamts génère un numéro « n0 » d'anonymat à partir du Nir de l'assuré pour chaque assuré sélectionné puis envoie la table de correspondance entre Nir de l'assuré et « n0 » aux services concernés des 3 caisses.

Chaque caisse envoie :

- un fichier contenant les coordonnées des assurés à contacter, référencé par le numéro « n0 », à la société chargée de l'enquête,
- le fichier de consommation de soins des assurés et de leur ayants droit, référencé par le numéro « n0 », à l'Insee.

La société d'enquête, après avoir réalisé les interviews, envoie les fichiers des données de l'enquête, référencés par le numéro « n0 », au Credes.

Après l'apurement des données, le Credes envoie les fichiers à l'Insee.

L'Insee génère à partir du numéro « n0 » un second numéro d'anonymat dit « n1 » sur les fichiers d'enquête reçus du Credes et sur les fichiers de consommation reçus de la CNAMTS, de la Canam et de la Msa. Il renvoie au Credes l'ensemble des fichiers de données anonymisées, identifiées par le numéro « n1 » du ménage. L'Insee sert donc uniquement de tiers responsable de la seconde anonymisation ; il n'apporte pas de données.

Le Credes procède à l'appariement des fichiers (le sexe, la date de naissance et les 4 premières lettres du prénom permettent de différencier les individus d'un même ménage).

Le numéro « n1 » est permanent au cours du temps pour un même foyer : il permet donc un chaînage longitudinal au cours des enquêtes successives (tous les 4 ans).

Documentation	Site internet : http://www.credes.fr Enquête sur la Santé et la protection sociale (ESPS) : méthode et déroulement en 2002 - Anne Doussin, Sylvie Dumesnil, Philippe Le Fur – Série « Méthode ». -Rapport n° 1 401 - Décembre 2002 Credes - Questions d'économie de la santé n°62 – décembre 2002
Contacts	Anne Doussin, Sylvie Dumesnil (Credes)

Maladies à déclaration obligatoire (MDO)

Les biologistes et les médecins ont l'obligation de déclarer à la Ddass 26 maladies à déclaration obligatoire.

En février 2002, la Direction générale de la santé a confié à l'Institut de veille sanitaire la mise en œuvre d'une procédure d'anonymisation des notifications des maladies à déclaration obligatoire, prévue par la loi du 1^{er} juillet 1998 relative au renforcement de la veille sanitaire (décret d'application du 6 mai 1999, modifié par le décret du 16 mai 2001).

La procédure d'anonymisation (développée par une société de services) est réalisée à l'aide d'une technique de hachage. Les identifiants (initiale du nom, prénom, sexe et date de naissance) sont transformés en un code anonyme de 16 caractères. Ces identifiants ont été choisis pour être compatibles avec les données antérieurement recueillies. Cette anonymisation est réalisée par les biologistes et les médecins pour l'infection à Vih, le Sida et l'hépatite B, par les Ddass pour toutes les autres maladies.

Le logiciel devrait être implanté dans les laboratoires et les hôpitaux selon le scénario suivant : le laboratoire qui réalise la sérologie crée le code et l'envoie au médecin qui complète les données médicales. L'InVS qui gère la base nationale, procède à une seconde anonymisation avec une clé secrète.

Fin 2002, un test des fonctionnalités du logiciel était en cours de validation.

Contacts	Contact téléphonique avec Christine Saura (InVS)
Textes officiels	Loi du 1 ^{er} juillet 1998 relative au renforcement de la veille sanitaire et décret d'application du 6 mai 1999 modifié par le décret du 16 mai 2001 Délibération Cnil n°02-082 du 19 novembre 2002

Au total, il existe d'ores et déjà en France des systèmes d'anonymisation et de chaînage, agréés par la Cnil, intégrés à des applications souvent d'envergure nationale. Un dispositif du même type pourrait être adopté pour apparier les bases hospitalières du Pmsi à une base de mortalité.

La faisabilité technique, institutionnelle et réglementaire d'un tel projet fera l'objet d'un prochain rapport.

Tableau récapitulatif des principaux dispositifs d'appariement de données développés en France

Organisme	DIM CHU Dijon	CESSI/CNAMTS	INVS (société de service)	INSERM (EMI 0214)	CREDES
Date début	1995	1996	2002 (phase test)	1992	1988
Nom	ANONYMAT	FOIN	ND	algorithme de San Marco	ND
Technique	Hachage (« Standard Hash Algorithm »)		Hachage	Hachage	
Identifiants	<ul style="list-style-type: none"> - nom - nom de jeune fille - prénom - DN - etc. 	<ul style="list-style-type: none"> - N° Sécurité sociale ouvrant droit - sexe bénéficiaire - DN bénéficiaire 	<ul style="list-style-type: none"> - initiale du nom - prénom - sexe - date de naissance 	<ul style="list-style-type: none"> - nom de naissance - prénom - jour et mois DN 	<ul style="list-style-type: none"> - N° Sécurité Sociale ouvrant droit
Chaînage	Probabiliste	Direct	direct	direct	direct
Applications	<ul style="list-style-type: none"> - Inter file active oncologie Rhône Alpes - Réseau périnatal et hépatite C en Bourgogne - Hépatite C et Suicides en Franche Comté - Collaboration équipe Cnrs génétique Grenoble - Collaborations Suisse et Luxembourg 	<ul style="list-style-type: none"> - Pmsi privé et public - Sniir-Am - Observatoire RMI - Études diverses 	Déclaration des MDO	DMI2	ESPS

F. La mortalité comme indicateur d'évaluation de la qualité des soins

L'utilisation de la mortalité comme indicateur de performance hospitalière présente des difficultés méthodologiques qui ont été largement étudiées.

En effet, des différences entre taux de mortalité peuvent refléter des différences dans la qualité de la prise en charge, mais aussi ^[5] :

- la politique d'admission et/ou de sortie de l'établissement ^{[40] [144]},
- la manière dont les facteurs de risque ont été pris en compte par le modèle d'ajustement,
- la qualité et l'exhaustivité du codage des variables utilisées pour l'ajustement,
- un facteur résiduel lié au hasard.

En théorie, pour pouvoir imputer des différences de mortalité à des différences de prise en charge, il faudrait que les patients aient, indépendamment de la prise en charge considérée, le même risque de décéder. Seule une randomisation des patients dans le cadre d'un essai clinique permettrait de se trouver dans ces conditions. À défaut, c'est à dire en situation d'observation comme c'est le cas en évaluation, il faut s'efforcer de rendre le plus comparable possible les groupes que l'on veut comparer, en minimisant le rôle de facteurs liés au pronostic vital indépendants de la qualité des soins. Il faut en outre disposer de la puissance statistique nécessaire pour pouvoir interpréter les résultats.

F.1 Contraintes méthodologiques

Pour pouvoir comparer des taux de mortalité, il faut :

1. **Définir la population cible** : on privilégiera les comparaisons de groupes de patients atteints d'une même pathologie ou ayant subi une même intervention chirurgicale. En effet, les comparaisons globales par hôpital portent sur des populations de patients et des pratiques beaucoup trop hétérogènes en terme de risque de mortalité ou de qualité des soins pour être informatives. Il a été montré que les indicateurs de mortalité, utilisés comme indicateurs de performance, peuvent différer au sein d'un même hôpital en fonction de la pathologie étudiée ^[120].
2. **Définir les critères diagnostiques** de la pathologie étudiée : ils doivent être explicites afin d'homogénéiser la définition des « cas ». La sévérité des critères diagnostiques utilisés aura en effet une influence sur les taux de mortalité observés. Si les critères diagnostiques ne sont pas clairement définis, les patients inclus peuvent avoir des pronostics vitaux différents et ce biais de sélection pourrait fausser la comparabilité entre les services ^{[24] [74] [86]}.
3. **Ajuster sur les facteurs de risque** : les populations de malades traités peuvent être très différentes d'un hôpital à l'autre, selon l'implantation, la politique de recrutement ou le type d'établissement. Comparer des taux bruts n'a que peu d'intérêt car les différences de mortalité observées risquent d'être surtout le reflet des différences de case-mix. Pour réduire au maximum ces facteurs de confusion, il est donc indispensable de travailler sur des taux ajustés ^{[1] [5] [14] [15] [23] [24] [29] [37] [40] [44] [47] [57] [58] [60] [62] [77] [82] [85] [100] [123] [129] [138] [145]}.

Les principaux domaines d'ajustement concernent :

- les caractéristiques socio-démographiques du patient,
- les critères de sévérité de la maladie et les comorbidités,
- l'aspect urgent ou programmé de la prise en charge,
- le type d'établissement (activité, taille, localisation etc..).

Le choix des variables pour prendre en compte ces domaines dépend de la pathologie considérée et parfois du pays. Un équilibre doit être trouvé sur le nombre de variables à inclure (trop faible, il risque d'aboutir à des conclusions erronées mais trop élevé, il nuit à l'intelligibilité des résultats et peut entraîner des problèmes d'estimation statistique). Les résultats comparatifs peuvent différer selon le nombre et la nature des facteurs de risque pris en compte et le choix du modèle d'ajustement est important ^[75].

Les modèles d'ajustement nécessitent souvent des variables non collectées en routine dans les systèmes d'information hospitaliers de type médico-technique (Pmsi ou DRG) et ne peuvent être réalisés qu'à partir de nouveaux systèmes d'information, spécifiques de la problématique étudiée ^{[47][71]}.

Idéalement, le modèle d'ajustement devrait être spécifique de la pathologie étudiée et avoir un bon pouvoir prédictif de la mortalité. Il ne devrait être utilisé que sur des populations identiques à celles ayant permis sa validation ^[12].

4. Disposer du nombre d'évènements nécessaire

Pour pouvoir interpréter des différences entre taux de mortalité, il faut disposer d'un nombre d'évènements suffisant. Ce problème de puissance statistique limite l'utilisation de l'indicateur aux structures ayant une forte activité, et à des pathologies ou des interventions responsables d'une mortalité élevée. À défaut, il est nécessaire de travailler à un niveau d'analyse plus agrégé. La comparaison peut être rendue difficile si le nombre d'évènements est très différent d'une structure à l'autre ^{[5][51][77][81][100][102].[139]}.

F.2 Qualité des données

Quelles que soient les précautions méthodologiques observées, la qualité des données utilisées est bien entendu, déterminante. Il faut s'assurer :

1. de la **qualité de l'information « décès »** : il faut s'assurer de l'exhaustivité de l'enregistrement des décès qui ne peut être assurée, dès lors que l'on travaille sur un grand nombre de cas, que par un système de chaînage automatisé et fiable. Seuls les pays disposant d'un tel système sont capables de produire des résultats à grande échelle et de participer à des comparaisons internationales ^{[82][103]}.
2. de la **qualité du recueil des variables d'ajustement** : les données doivent être homogènes en termes d'exhaustivité et de qualité de codage. Une sous-déclaration des comorbidités ou des erreurs de codage du diagnostic principal par exemple mettent en cause la comparabilité des taux de mortalité. Si les informations sont issues du système d'information hospitalier, elles sont par définition standardisées ; par contre la qualité du codage peut varier d'un hôpital à l'autre et au cours du temps (faussant alors les

comparaisons chronologiques) [5] [16] [22] [45] [59] [78]. Si le recueil des données est réalisé pour une étude ad hoc, il faut s'assurer de l'utilisation d'un questionnaire standardisé.

La mise en place d'indicateurs de mortalité devrait toujours être accompagnée d'une évaluation du codage pour minimiser ce biais.

F3. Définition de la période d'observation

L'indicateur peut porter sur la mortalité **intra-hospitalière** ou **post-hospitalière**.

La mortalité **intra-hospitalière** a longtemps été utilisée comme la mesure standard [64] jusqu'à ce que l'étude de Jencks [62] montre qu'une période d'observation constante était préférable au temps d'observation hospitalier dont la durée varie avec les comorbidités et d'autres facteurs influençant la durée de séjour. La politique de l'hôpital ou du service en matière d'admission et de sortie, la stratégie de recrutement, l'existence ou non de services de réanimation et de soins palliatifs, l'organisation de la filière d'aval conditionnent à la fois le profil des malades pris en charge et la durée de séjour à l'hôpital [40] [82] [101] [144]. En outre, la durée moyenne de séjour évolue dans le temps. Goldacre par exemple a étudié l'évolution en Grande-Bretagne de la proportion intra-hospitalière des décès post-chirurgie à 30 jours. Sur une période de 35 ans (de 1963 à 1998) la part intra-hospitalière des décès (n=41 200) est passée de 79 % pour la période 1963-1974 à 61 % pour 1987-98. Cette évolution s'explique par la diminution concomitante de la durée de séjour [42].

Les comparaisons de taux de mortalité intra-hospitalière peuvent donc conduire à des conclusions erronées [3] [65] [143].

L'utilisation de la mortalité **post-hospitalière** permet d'éviter les biais qui viennent d'être cités. Les définitions de la mortalité post-hospitalière utilisées par les différents auteurs varient à la fois par la durée de suivi (30, 90, 180 jours, etc.), par l'événement datant le début de l'observation (admission, sortie, procédure) et par la définition des cas à inclure [64].

Quel délai d'observation choisir ?

L'objectif est de choisir la période la plus courte possible qui inclut la plus grande partie des décès. Une période trop courte sous-estimerait le nombre de décès liés à la pathologie étudiée ; à l'inverse, plus le délai est long, plus il est difficile d'attribuer le décès à la pathologie ou à l'intervention étudiée, en particulier chez les personnes âgées [82]. Cette période varie selon le type de pathologie. Seagroatt par exemple a étudié l'évolution de la mortalité au cours de la première année suivant une intervention chirurgicale et montre qu'elle est variable selon le type d'intervention : le pic de mortalité précoce (<30 jours) observé pour certaines interventions ne l'est pas pour d'autres [123]. Le délai d'observation choisi pourrait modifier les résultats comparatifs entre structures [33] [64].

D'après les chiffres obtenus auprès du CépiDc pour l'année 1997, la part des décès survenus à l'hôpital ou dans une clinique privée en France était de 58 %. Ce chiffre variait avec le sexe (60 % pour les hommes versus 56 % pour les femmes) et l'âge (37 % pour les garçons de 15-24 ans à 64 % pour les hommes de 65-84 ans par exemple).

F.4 Conclusion

Au total, la mortalité est un indicateur de résultat délicat à manier dès lors que l'on souhaite l'utiliser pour évaluer la qualité des soins. La relation entre taux de mortalité et qualité des soins est controversée ^{[5] [28] [51] [72] [93] [100] [103] [111] [137] [141] [142]}.

Même si certains biais liés à une période d'observation purement hospitalière sont minimisés par l'utilisation de la mortalité post-hospitalière, le choix du modèle d'ajustement, la mise à disposition des données nécessaires, les contraintes de puissance statistique, la nécessité d'homogénéiser la qualité du recueil et du codage des données posent des problèmes logistiques et de coût non négligeables. En outre, malgré de nombreuses améliorations, les méthodologies d'ajustement sur les facteurs de risque sont loin d'être parfaites. Ces contraintes méthodologiques devraient limiter son utilisation à certaines pathologies à forte létalité, pour lesquelles des procédures de soins sont bien standardisées (comme par exemple le champ de la chirurgie cardiaque).

Malgré toutes ces réserves, elle peut apporter des informations utiles à condition de l'utiliser dans des conditions bien maîtrisées, en définissant clairement les limites de son interprétation. Les résultats de mortalité peuvent notamment servir « d'événement sentinelle » incitant à une réflexion sur les pratiques et non de méthode de contrôle ^{[1] [39] [75] [85]}.

D'autres indicateurs, notamment d'organisation et de procédures, doivent y être associés pour éclairer les chiffres de mortalité et aider à la mise en place de mesures de correction le cas échéant ^{[49] [148]}.

Enfin, comme pour tout indicateur, le niveau d'implication des professionnels de santé lors de sa mise en place, la crédibilité technique de l'outil (qui devrait être calculé de préférence par un organisme indépendant) et les modalités de restitution des résultats conditionnent son impact.

F.5 Impact de la publication d'un indicateur de mortalité

Des résultats comparatifs de taux de mortalité sont diffusés dans plusieurs pays (États-Unis, Écosse, Angleterre, etc ^{[24] [84] [146]}).

En France, des résultats de performance hospitalière incluant des taux de mortalité ont commencé à être publiés par les médias à la fin des années 90 ^[54].

Le contexte de ces publications (organismes à l'initiative de la diffusion, objectifs affichés, contexte sanitaire et économique), ainsi que le mode de présentation des résultats (niveau d'agrégation des résultats, « classement » ou non des établissements etc.) varient selon les pays.

L'intérêt de la diffusion publique d'indicateurs de résultats est très discuté.
(^{[3] [4] [13] [18] [19] [27] [28] [31] [39] [45] [48] [77] [82] [84] [85] [88] [103] [110] [111] [113] [122] [141] [147] [148]})

Certains auteurs remettent en question la valeur de l'indicateur : une forte mortalité ne pourrait être imputée à une mauvaise qualité des soins, alors que tous les facteurs influents ne peuvent être contrôlés. Ils considèrent donc que les classements ne reflètent pas de réelles différences de qualité entre établissements et que leur diffusion constitue une fausse information.

Pour d'autres au contraire, la publication des résultats permettrait de :

- favoriser le dialogue entre les cliniciens et la direction,
- stimuler la recherche pour comprendre les différences de résultats et mettre en place des mesures d'amélioration (réorganisation du service, formation permanente, recrutement de personnel spécialisé etc.),
- produire une connaissance utile pour l'amélioration des performances,
- repérer des problèmes de qualité de recueil ou de codage des données.

D'autres enfin craignent que cette publication :

- modifie le recrutement des patients (par exemple, refus de chirurgiens d'opérer les patients à haut risque,
- modifie les choix thérapeutiques,
- entraîne une sur-déclaration des comorbidités et de la sévérité (afin d'améliorer les résultats),
- focalise l'attention des professionnels sur les seules pathologies pour lesquelles des indicateurs sont calculés,
- réassure à tort des services dont les résultats sont « dans la moyenne »,
- inquiète le public s'il n'a pas le choix entre plusieurs structures dans sa région.

Sur un plan pratique enfin, les délais de diffusion des résultats sont souvent perçus comme trop longs (des modifications de pratiques ayant pu intervenir entre-temps).

Enfin, les résultats de mortalité sont souvent considérés comme difficiles à utiliser pour améliorer les pratiques s'ils ne sont pas associés à d'autres types d'information.

Impact de la publication d'un indicateur de mortalité

USA

Dziuban ^[31] - Un groupe de travail multidisciplinaire a entrepris une revue des dossiers pour essayer de comprendre pourquoi la mortalité après pontage coronarien était plus élevée que la moyenne à l'hôpital de St Peter (Albanie, État de New York). La surmortalité concernait un petit groupe de patients à haut risque. Un programme d'amélioration de la prise en charge de ces patients a pu être mis en place et une baisse de la mortalité pour le groupe à haut risque et pour les autres a été observée depuis 1993. L'auteur conclut que l'utilisation d'un seul chiffre de performance peut cacher plus d'information qu'il n'en révèle et que seul un examen approfondi de la base de données permet d'identifier les domaines susceptibles d'être améliorés.

Burack ^[13] - Enquête sur l'impact de la diffusion des résultats auprès des 150 chirurgiens cardiaques de New York à l'aide d'un questionnaire anonyme (69,3 % de réponse) : 70 % n'ont pas modifié leur pratique. Beaucoup ont pointé la mauvaise définition de la maladie pulmonaire obstructive (45 %) ou des méthodes statistiques (60 %). 62 % ont refusé d'opérer au moins un patient à haut risque dans l'année, surtout à cause de la diffusion publique des résultats. Le refus concernait plus les chirurgiens ayant moins d'expérience (< 10 ans ou < 100 actes ou non spécialistes en chirurgie cardiothoracique). Plus de patients étaient traités non chirurgicalement.

Hannan ^[48] - Un questionnaire a été envoyé à tous les cardiologues de l'État de New York (450 réponses) : 22 % discutent des résultats diffusés avec leurs patients et 38 % déclarent que l'information a modifié leur correspondants chirurgiens un peu ou beaucoup.

Schneider ^[122] - Les taux de mortalité post pontage coronarien sont diffusés en Pennsylvanie depuis 1992 par hôpitaux et par chirurgiens. Une enquête a été réalisée en 1995 auprès d'un échantillon représentatif de 50 % des cardiologues et des chirurgiens cardiaques : 82 % des cardiologues et tous les chirurgiens connaissaient le guide. Seulement 10 % considéraient le taux de mortalité comme très important pour juger un chirurgien. Moins de 10 % en parlaient avec plus de 10 % de leurs patients candidats à 1 pontage. 87 % des cardiologues déclaraient que le guide avait pas ou très peu d'influence sur leur décision d'adresser un patient à un chirurgien. Les critiques les plus fréquentes portaient sur l'absence indicateurs de qualité autres que la mortalité (78 %), un ajustement inadéquat (79 %) et le manque de fiabilité des données transmises par les hôpitaux et les chirurgiens (53 %). 59 % des cardiologues rapportaient des difficultés croissantes pour trouver des chirurgiens acceptant de réaliser un pontage à des patients sévèrement malades et 63 % des chirurgiens déclaraient opérer moins volontiers de tels patients.

Rainwater ^[110] - Enquête en Californie auprès de la direction des hôpitaux : la diffusion des résultats a peu d'impact sur les procédures de soins pour les patients avec infarctus du myocarde. 2 critiques essentielles sont formulées : le délai de diffusion (trop long) et l'absence d'information jointe pour aider à améliorer les procédures de soins.

Bentley ^[4] - Enquête auprès des hôpitaux, des payeurs et des acheteurs en Pennsylvanie (où les résultats post pontage coronarien étaient publiés annuellement) et au New Jersey (où il n'y avait pas de diffusion de résultats). Des modifications dans les domaines du marketing, de la gestion et des soins ont été observées en Pennsylvanie (par exemple 38 % des hôpitaux utilisent les résultats de performance pour recruter des chirurgiens spécialisés), ce qui n'était pas le cas au New Jersey.

L'exemple de l'Ecosse ^[20]

Depuis 1993 le Clinical Resource and Audit Group (CRAG) du NHS publie en Écosse des indicateurs de résultats cliniques comparatifs (NB : ensemble d'indicateurs ne se limitant pas au domaine hospitalier ni à la mortalité). En 2000, une équipe a été chargée d'évaluer l'impact de la diffusion de ces indicateurs. Le recueil des données s'est fait à l'aide d'interviews, de réunions avec les différents professionnels de la santé, d'une recherche bibliographique et d'observations recueillies lors de meetings.

Les réactions enregistrées rejoignent les principaux arguments déjà listés page précédente.

Conclusions du groupe en charge de l'enquête :

- il faudrait plus de discussions avec les cliniciens avant de publier,
- il ne faut pas utiliser les indicateurs dans une attitude de confrontation mais de dialogue,
- les rapports devraient être plus dynamiques, plus instructifs. Des informations plus utiles au niveau local, et des plans d'action devraient être fournis.
- l'impact de la diffusion publique des rapports a évolué: elle suscite maintenant beaucoup moins d'intérêt qu'au début, elle est traitée de manière moins sensationnelle et les hôpitaux la ressentent de manière moins défensive. Elle risque pourtant parfois de démoraliser les équipes et de provoquer de l'anxiété chez les patients qui n'ont pas le choix entre plusieurs structures.

G. Conclusion et perspectives

Au delà des discussions méthodologiques sur la capacité de l'indicateur « mortalité » à déterminer la performance d'un établissement par rapport à un autre, il n'en demeure pas moins que la connaissance du statut vital d'un individu, suite à sa prise en charge hospitalière, constitue un indicateur majeur et incontournable en terme de santé publique.

La crainte habituelle des professionnels de santé et particulièrement des médecins est l'utilisation malgré toutes les réserves méthodologiques décrites de cet indicateur à des fins de comparaison inter hospitalière. Ces réserves, si elles ne sont pas dénuées de fondement, ne doivent pas pour autant masquer l'avancée en terme d'amélioration des connaissances épidémiologiques que constituerait la mise à disposition des mortalités post hospitalières par type de prise en charge.

Cette problématique s'inscrit aujourd'hui dans un contexte à la fois technique et politique favorable.

D'un point de vue technique, grâce à la mise en place du Pmsi, les avancées substantielles depuis ces dix dernières années des systèmes d'information des établissements hospitaliers de court séjour est une réalité. La mise en cohérence du Pmsi public et privé permet de disposer au niveau régional de bases de données identiques, chaînables depuis 2001 toutes ensembles grâce à la procédure Foin.

Du point de vue des politiques hospitalières, la réforme de la Tarification à l'activité (TAA) met en exergue la nécessité de disposer d'indicateurs « qualité » capables d'établir si la mise en place de réformes du financement du secteur hospitalier peut avoir un impact négatif sur la qualité des pratiques médicales. En dehors des indicateurs habituels, tels que le nombre de réadmissions précoces ou de transferts, capables de mettre à jour des modifications de pratiques perverses, la mortalité décalée apparaît comme une information épidémiologique centrale capable d'éclairer cette problématique.

Cependant, l'intérêt d'un tel indicateur dans le paysage sanitaire national ne doit pas être dévoyé par une utilisation inadéquate, tant dans ses modalités de mise en place que de restitution, qui rendrait difficile son appropriation par le corps médical. En effet, l'utilité d'un indicateur quel qu'il soit, n'est pas lié uniquement à ses qualités intrinsèques. Les modalités organisationnelles de sa mise en place et de son utilisation conditionnent son impact.

Il convient donc de définir précisément le champ d'utilisation de cet indicateur et surtout de s'interdire toute utilisation scientifiquement incorrecte, telle que le classement d'établissements. Ce type de dérive serait particulièrement contre-productive, même si par ailleurs la demande d'indicateur de qualité hospitalier est et sera de plus en plus prégnante.

Un des moyens de répondre aux problèmes liés au manque de puissance des tests disponibles est d'agréger l'information à un niveau supra hospitalier, ce qui permet en outre de ne pas stigmatiser les établissements. Pour autant, on ne peut pas exclure de la réflexion l'utilisation de ces résultats à un niveau local, non pour critiquer la qualité de la prise en charge médicale d'un établissement, mais pour engager une réflexion sur les éventuelles améliorations organisationnelles à apporter en amont (soins primaires, prise en charge des urgences, organisation des transports sanitaires...) et en aval (filrière de soins, HAD...) de l'hôpital.

Cette réflexion concerne non seulement les établissements en interne mais également les interactions entre les différents types de structures. Elle doit être intégrée aux actions de planification et de régulation menées par l'Agence régionale d'hospitalisation.

Au niveau national, les comparaisons inter régionales permettraient, en ciblant de fortes disparités, d'établir des programmes d'actions prioritaires spécifiques à chaque région en fonction des choix des acteurs locaux et de suivre l'évolution de ces résultats. En effet, la pérennité d'un tel système d'information confère à l'indicateur la capacité d'évaluer dans le temps l'effet des actions menées.

Par ailleurs, une base de mortalité anonyme et « chaînable » avec d'autres bases de données pourrait être utile dans le cadre d'études scientifiques nécessitant la mise à disposition du statut vital des individus : études épidémiologiques, recherche clinique, suivi de registres, évaluation des pratiques, recherche en santé publique, etc.

Au niveau international enfin, la France pourrait participer aux études menées par les organismes européens ou internationaux (OCDE, OMS...) destinées à comparer les performances globales des systèmes sanitaires en terme de mortalité « décalée » pour des pathologies ou des actes techniques spécifiques.

H. Références

- [1] Ansari MZ , Ackland MJ. Inter-hospital comparison of mortality rates. *Int J Qual Health Care* 1999 Feb ;11(1):29-35
- [2] Arellano MG, Weber GI. Issues in identification and linkage of patient records across an integrated delivery system. *J Health Inf Manag* 1998 Fall ;12(3) : 43-52
- [3] Baker DW, Einstadter D, Thomas CL. Mortality trends during a program that publicly reported hospital performance. *Med Care* 2002 Oct;40(10):879-90
- [4] Bentley JM, Nash DB. How Pennsylvania hospitals have responded to publicly released reports on coronary artery bypass graft surgery. *Jt Comm J Qual Improv* 1998 Jan ; 24(1) : 40-9
- [5] Best WR, Cowper DC. The ratio of observed-to-expected mortality as a quality of care indicator in non-surgical VA patients. *Med Care* 1994 Apr ; 32(4) : 390-400
- [6] Black CD, Burchill CA, Roos LL. The Population Health Information System: data analysis and software. *Med Care* 1995 Dec ; 33(12 Suppl) : DS127-31
- [7] Blomqvist P, Ekblom A, Nyren O, Krusemo UB, Bergstrom R, Adami HO. Survival after rectal cancer: differences between hospital catchment areas. A nationwide study in Sweden. *Gut* 1999 Jul ; 45(1) : 39-44
- [8] Borst F, Allaert FA, Quantin C. The Swiss solution for anonymously chaining patient files. *Medinfo* 2001 ; 10(Pt 2) : 1239-41
- [9] Brameld K, Holman D, Thomas M, Bass J. Use of a state data bank to measure incidence and prevalence of a chronic disease: end-stage renal failure *Am J Kidney Dis* 1999 Dec ; 34(6) : 1033-9
- [10] Brameld KJ, Holman CD, Bass AJ, Codde JP, Rouse IL. Hospitalisation of the elderly during the last year of life: an application of record linkage in Western Australia 1985-1994. *J Epidemiol Community Health* 1998 Nov ; 52(11) : 740-4
- [11] Brameld KJ, Holman CD, Threlfall TJ, Lawrence DM, De Klerk NH. Increasing 'active prevalence' of cancer in Western Australia and its implications for health services. *Aust N Z J Public Health* 2002 Apr ; 26(2) : 164-9
- [12] Bridgewater B, Neve H, Moat N, Hooper T, Jones M. Predicting operative risk for coronary artery surgery in the United Kingdom: a comparison of various risk prediction algorithms. *Heart* 1998 Apr ; 79(4) : 350-5
- [13] Burack JH, Impellizzeri P, Homel P, Cunningham JN Jr. Public reporting of surgical mortality: a survey of New York State cardiothoracic surgeons. *Ann Thorac Surg* 1999 Oct;68(4) : 1195-200 ; discussion 1201-2
- [14] Capewell S, Kendrick S, Boyd J, Cohen G, Juszczak E, Clarke J. Measuring outcomes: one month survival after acute myocardial infarction in Scotland. *Heart* 1996 Jul;76(1):70-5
- [15] Capewell S, MacIntyre K, Stewart S, Chalmers JW, Boyd J, Finlayson A, Redpath A, Pell JP, McMurray JJ. Age, sex, and social trends in out-of-hospital cardiac deaths in Scotland 1986-95: a retrospective cohort study. *Lancet* 2001 Oct 13 ; 358(9289) : 1213-7
- [16] Cauvin JM, Hardy B, Jehan P et al. Qualité du codage et conséquences en ISA et journées théoriques : à partir de 100 dossiers recodés par établissement. *J économie médicale* 1997 ; 15 ; 3-4 : 195-206
- [17] Chamberlayne R, Green B, Barer ML, Hertzman C, Lawrence WJ, Sheps SB. Creating a population-based linked health database: a new resource for health services research. *Can J Public Health* 1998 Jul-Aug ; 89(4) : 270-3

- [18] Chassin MR, Hannan EL, DeBuono BA. Benefits and hazards of reporting medical outcomes publicly. *N Engl J Med* 1996 Feb 8 ; 334(6) : 394-8
- [19] Chassin MR. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. *Health Aff (Millwood)* 2002 Jul-Aug ; 21(4) : 40-51
- [20] Clinical Outcome Indicators, Clinical Outcomes Working Group. The Scottish Executive (may 2002)
- [21] CNAM-TS/DSI/CESSI. « Fonction d'occultation d'identifiants nominatifs ». Mai 2002
- [22] D'alche-Gazutier M-J, Hamon-Poupinel V, Henriet L et al. La qualité des informations médicales du PMSI : comparaison de deux enquêtes à deux ans d'intervalle. *J économie médicale* 1999 ; 17 ; 1 : 29-36
- [23] Davies HT, Crombie IK. Assessing the quality of care. *BMJ* 1995 Sep 23;311(7008):766
- [24] Dillner L. Scottish death rates published with health warning. *BMJ* 1994 Dec 17 ; 309(6969) : 1599-600
- [26] DRESS 2001. Éléments pour évaluer la performance des établissements hospitaliers. *Dossiers solidarité et santé* n° 2
- [27] DREES 2001. L'information de l'usager/consommateur sur la performance du système de soins – *Revue bibliographique. Études* n° 13
- [28] De Pourville G, Minvielle E. « La mesure de la qualité des soins à l'hôpital : l'état de l'art. Quelle information donner au public ? ». OCDE 2002
- [29] Dubois RW, Brook RH, Rogers WH. Adjusted hospital death rates: a potential screen for quality of medical care. *Am J Public Health* 1987 Sep ; 77(9) : 1162-6
- [30] Dusserre L, Quantin C., Bouzelat H. A one way public key cryptosystem for the linkage of nominal files in epidemiological studies. *Medinfo* 1995 ; 8 Pt 1 : 644-7
- [31] Dziuban SW Jr. Using information from databases to improve clinical practice: lessons learned under fire. *Ann Thorac Surg* 1997 Dec ; 64(6 Suppl) : S64-7 ; discussion S80-2
- [32] Fair M, Cyr M, Allen AC, Wen SW, Guyon G, MacDonald RC. An assessment of the validity of a computer system for probabilistic record linkage of birth and infant death records in Canada. The Fetal and Infant Health Study Group. *Chronic Dis Can* 2000 ; 21(1) : 8-13
- [33] Garnick DW. Measuring hospital mortality rates: are 30-day data enough? Ischemic Heart Disease Patient Outcomes Research Team. *Health Serv Res* 1995 Feb ; 29(6) : 679-95
- [34] Gentot M. La protection des données à caractère personnel : un impératif pour la recherche épidémiologique. *Rev Epidemiol Sante Publique* 2000 Jan ; 48(1) : 3-5
- [35] Gill L, Goldacre M, Simmons H, Bettley G, Griffith M. Computerised linking of medical records: methodological guidelines. *J Epidemiol Community Health* 1993 Aug ; 47(4) : 316-9
- [37] Glance LG, Osler TM. Comparing outcomes of coronary artery bypass surgery: Is the New York Cardiac Surgery Reporting System model sensitive to changes in case mix? *Crit Care Med* 2001 Nov ; 29(11) : 2090-6
- [38] GMSIH (Groupement pour la Modernisation de Système d'Information Hospitalier). Principes et processus d'identification du patient – Synthèse de l'analyse réglementaire. Mars 2001
- [39] Goddard M, Davies HT, Dawson D, Mannion R, McInnes F. Clinical performance measurement: part 2- avoiding the pitfalls. *J R Soc Med* 2002 Nov ; 95(11) : 549-51
- [40] Goldacre M, Gill L. Interpreting hospital death rates. *BMJ* 1995 ; 310 : 599(4 march)

- [41] Goldacre M, Kurina L, Yeates D, Seagroatt V, Gill L. Use of large medical databases to study associations between diseases. *QJM* 2000 Oct ; 93(10) : 669-75
- [42] Goldacre MJ, Griffith M, Leicester G, Mackintosh A. In-hospital deaths as fraction of all deaths within 30 days of hospital admission for surgery : analysis of routine statistics. *BMJ*. 2002 May 4 ; 324(7345) : 1069-70
- [43] Goldacre MJ, Simmons H, Henderson J, Gill LE. Trends in episode based and person based rates of admission to hospital in the Oxford record linkage study area. *Br Med J (Clin Res Ed)* 1988 Feb 20 ; 296(6621) : 583-5
- [44] Green J, Passman LJ, Wintfeld N. Analyzing hospital mortality. The consequences of diversity in patient mix. *JAMA* 1991 Apr 10 ; 265(14) : 849-53
- [45] Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? *Med Care* 1993 Aug ; 31(8) : 719-31
- [46] Hammar N, Alfredsson L, Rosen M, Spetz CL, Kahan T, Ysberg AS. A national record linkage to study acute myocardial infarction incidence and case fatality in Sweden. *Int J Epidemiol* 2001 Oct;30 Suppl 1 : S30-4
- [47] Hannan EL, Kilburn H Jr, Lindsey ML, Lewis R. Clinical versus administrative data bases for CABG surgery. Does it matter? *Med Care* 1992 Oct ; 30(10) : 892-907
- [48] Hannan EL, Stone CC, Biddle TL, DeBuono BA. Public release of cardiac surgery outcomes data in New York: what do New York state cardiologists think of it? *Am Heart J* 1997 Dec ; 134(6) : 1120-8
- [49] Hannan EL, Magaziner J, Wang JJ. Mortality and locomotion 6 months after hospitalization for hip fracture: risk factors and risk-adjusted hospital outcomes. *JAMA* 2001 Jun 6 ; 285(21) : 2736-4
- [50] Hargreave TB, Heynes CF, Kendrick SW, Whyte B, Clarke JA. Mortality after transurethral and open prostatectomy in Scotland. *Br J Urol* 1996 Apr ; 77(4) : 547-53
- [51] Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Med Care* 1996 Aug ; 34(8) : 737-53
- [52] Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust N Z J Public Health* 1999 Oct ; 23(5) : 453-9
- [53] Holman CD, Wisniewski ZS, Semmens JB, Rouse IL, Bass AJ. Mortality and prostate cancer risk in 19,598 men after surgery for benign prostatic hyperplasia. *BJU Int* 1999 Jul ; 84(1) : 37-42
- [54] Houdart Ph, Malye F, Vincent J. Le palmarès des hôpitaux. *Sciences et Avenir* Sept 1998 ; 619 : 32-71
- [55] Howell SJ, Sear JW, Sear YM, Yeates D, Goldacre M, Foex P. Risk factors for cardiovascular death within 30 days after anaesthesia and urgent or emergency surgery: a nested case-control study. *Br J Anaesth* 1999 May ; 82(5) : 679-84
- [56] Howell SJ, Sear YM, Yeates D, Goldacre M, Sear JW, Foex P. Risk factors for cardiovascular death after elective surgery under general anaesthesia. *Br J Anaesth* 1998 Jan ; 80(1) : 14-9
- [57] Iezzoni LI. The risks of risk adjustment. *JAMA* 1997 Nov 19;278(19):1600-7
- [58] Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health* 1996 Oct;86(10):1379-87

- [59] Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA* 1992 Apr 22-29;267(16):2197-203
- [60] Jarman B. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999 Jun 5;318(7197):1515-2
- [61] Jaro MA. Probabilistic linkage of large public health data files *Stat Med* 1995 Mar 15-Apr 15;14(5-7):491-8
- [62] Jencks SF, Williams DK, Kay TL. Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. *JAMA* 1988 Oct 21;260(15):2240-6
- [63] Johansson LA, Westerling R. Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics. *Int J Epidemiol* 2000 Jun;29(3):495-502
- [64] Johnson ML, Gordon HS, Petersen NJ, Wray NP, Shroyer AL, Grover FL, Geraci JM. Effect of definition of mortality on hospital profiles. *Med Care* 2002 Jan;40(1):7-16
- [65] Kahn KL, Keeler EB, Sherwood MJ, Rogers WH, Draper D, Bentow SS, Reinisch EJ, Rubenstein LV, Kosecoff J, Brook RH. Comparing outcomes of care before and after implementation of the DRG-based prospective payment system. *JAMA* 1990 Oct 17;264(15):1984-8
- [66] Kelman CW, Bass AJ, Holman CD. Research use of linked health data--a best practice protocol. *Aust N Z J Public Health* 2002;26(3):251-5
- [67] Kendrick SW. The développement of record linkage in Scotland : the responsive application of probability matching. *Record Linkage Techniques 1997 - Proceedings International Workshop and Exposition - Arlington, VA March 20-21, 1997*
- [68] Kendrick SW, Douglas MM, Gardner D, Hucker D. Best-link matching of Scottish health data sets. *Methods Inf Med* 1998 Jan;37(1):64-8
- [69] Kerkri EM, Quantin C, Allaert FA, Cottin Y, Charve P, Jouanot F, Yetongnon K. An approach for integrating heterogeneous information sources in a medical data warehouse. *J Med Syst* 2001 Jun;25(3):167-76
- [70] M. Kerkri, C. Quantin, T. Grison, K. Yetongnon. A methodology for multi-source record linkage. *The XVIIIth international Congress of the European Federation for Medical Informatics, St Malo, France, du 4 au 7 mai 2003. A paraître.*
- [71] Krakauer H, Bailey RC, Skellan KJ, Stewart JD, Hartz AJ, Kuhn EM, Rimm AA. Evaluation of the HCFA model for the analysis of mortality following hospitalization. *Health Serv Res* 1992 Aug;27(3):317-35
- [72] Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented Internet health care report card: the risk of quality ratings based on mortality data. *JAMA* 2002 Mar 13;287(10):1277-87
- [73] Labillois T. A comparison of direct match and probabilistic linkage in the death clearance of the Canadian Cancer Registry. *Proceedings of an International Workshop and Exposition, Arlington, VA, March 20-21, 1997*
- [74] Lang T, Ducimetiere P, Arveiler D, Amouyel P, Ferrieres J, Ruidavets JB, Montaye M, Haas B, Bingham A. Trends and geographical disparities in coronary heart disease in France: are results concordant when different definitions of events are used? *Int J Epidemiol* 1999 Dec;28(6):1050-8
- [75] Lawrance RA, Dorsch MF, Sapsford RJ, Mackintosh AF, Greenwood DC, Jackson BM, Morrell C, Robinson MB, Hall AS. Use of cumulative mortality data in patients

- with acute myocardial infarction for early detection of variation in clinical practice: observational study. *BMJ* 2001 Aug 11;323(7308):324-7
- [76] Le Minous F, Jarno P, Ledoux B et al. Evaluation de la qualité du codage des séjours de cancérologie groupés dans le GHM 675. *J économie médicale* 2000;18;1:51-62
- [77] Localio AR, Hamory BH, Fisher AC, TenHave TR. The public release of hospital and physician mortality data in Pennsylvania. A case study. *Med Care* 1997 Mar;35(3):272-8
- [78] Lombraïl P, Minvielle E, Comar L, Gottot S. Prospective Payment Information Systems and epidemiology: a difficult link to establish. *Rev Epidemiol Sante Publique* 1994;42(4):334-44
- [79] Mahonen M, Salomaa V, Keskimaki I, Moltchanov V. The feasibility of routine mortality and morbidity register data linkage to study the occurrence of acute coronary heart disease events in Finland. *Eur J Epidemiol* 2000;16(8):701-1
- [80] Mahonen M, Salomaa V, Keskimaki I, Moltchanov V, Torppa J, Molarius A, Tuomilehto J, Sarti C. The feasibility of combining data from routine Hospital Discharge and Causes-of-Death Registers for epidemiological studies on stroke. *Eur J Epidemiol* 2000;16(9):815-7
- [81] Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995 Sep 23;311(7008):793-6
- [82] McKee M, Sanderson C. Commentary : measures of early postoperative mortality. *BMJ* 1994 August 6;309:365-6
- [83] McKee M. Indicators of clinical performance *BMJ* 1997;315:142
- [84] McKee M, Chenet L. Informing the market: the strengths and weaknesses of information in the British National Health Service. *Health Care Anal* 1997 Jun;5(2):149-5
- [85] McKee M, Hunter D. Mortality league tables: do they inform or mislead? *Qual Health Care* 1995 Mar;4(1):5-12
- [86] McKenna CJ and Forfar JC. What is heart attack ? Troponin positive acute coronary syndrome versus myocardial infarction. *BMJ*. 2002 Feb 16;324(7334):377-8
- [87] Melton LJ. History of the Rochester Epidemiology Project. *Mayo Clin Proc* 1996; 71:266-274
- [88] Mennemeyer ST, Morrissey MA, Howard LZ. Death and reputation: how consumers acted upon HCFA mortality information. *Inquiry* 1997 Summer;34(2):117-28
- [89] Moise P. Using hospital administrative databases for a disease-based approach to studying health care systems. *OECD ageing related disease study*. August 2001
- [90] Morrison H, Savitz D, Semenciw R, Hulka B, Mao Y, Morison D, Wigle D. Farming and prostate cancer mortality. *Am J Epidemiol* 1993 Feb 1;137(3):270-80
- [91] Morrison HI, Semenciw RM, Morison D, Magwood S, Mao Y. Brain cancer and farming in western Canada. *Neuroepidemiology* 1992;11(4-6):267-7
- [92] Naiditch M. L'exemple de la chirurgie cardiaque new-yorkaise. *La recherche* 324 oct 1999 : 66-9
- [93] Naiditch M., Durieux P., Lairy G. Les taux de mortalité hospitalière : intérêt et limites pour la mesure de la performance comparative des hôpitaux. *Solidarité et Santé* 2001;1:36-9
- [94] Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL, Wyse RK, Ferguson TB. Validation of European System for Cardiac Operative Risk Evaluation

- (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg* 2002 Jul;22(1):101-5
- [95] Newcombe HB. Strategy and art in automated death searches. *Am J Public Health* 1984 Dec;74(12):1302-3
- [96] Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic linkage of vital records. *Science* 130, 954-959 (1959)
- [97] Newcombe HB, Smith ME, Howe GR, Mingay J, Strugnell A, Abbatt JD. Reliability of computerized versus manual death searches in a study of the health of Eldorado uranium workers. *Comput Biol Med* 1983;13(3):157-69
- [98] O'Hara. Linking mortality data with hospital morbidity data - a pilot study. *Health information management* 1995;25(4):126-8
- [99] O'Hara. Mortality soon after discharge from a major teaching hospital : linking mortality and morbidity. *J. Qual. Clin. Practice* (1996) 16, 39-48
- [100] Park RE, Brook RH, Kosecoff J, Keeseey J, Rubenstein L, Keeler E, Kahn KL, Rogers WH, Chassin MR. Explaining variations in hospital death rates. Randomness, severity of illness, quality of care. *JAMA* 1990 Jul 25;264(4):484-9
- [101] Pates RD. Adding value to clinical data by linkage to a public death registry. *Medinfo* 2001;10(Pt 2):1384-8
- [102] Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *BMJ* 1998 Jun 6;316(7146):1697-70
- [103] Pouvourville G, Minvielle E. La mesure de la qualité des soins à l'hôpital : l'état de l'art. Quelle information donner au public ? Etre à la hauteur : mesurer et améliorer la performance des systèmes de santé dans les pays de l'OCDE - OCDE 2002
- [104] Proceedings International Workshop and Exposition -.Arlington, VA March 20-21, 1997. Horm J. Multiple Causes of death for the National Health Interview Survey.
- [105] Quantin C, Allaert F-A, Bouzelat H, Rodrigues J-M, Trombert-Paviot B, Brunet-Lecomte P, Grémy F, Dusserre L. Security of healthcare data networks used for epidemiological studies. *Rev Epidemiol Sante Publique* 2000 Jan;48(1):89-99
- [106] Quantin C, Kerkri E, Allaert F-A, Bouzelat H, Dusserre L. Security services: an overview of the French legislation on cryptography. *Rev Epidemiol Sante Publique* 2000 Jan;48(1):81-87
- [107] Quantin C, Allaert F-A, Dusserre L. Anonymous statistical methods versus cryptographic methods in epidemiology. *Int J Med Inf* 2000 Nov;60(2):177-83
- [108] Quantin C, Bouzelat H, Allaert F-A, Benhamiche A-M, Faivre J, Dusserre L. How to ensure data security of an epidemiological follow-up: quality assessment of an anonymous record linkage procedure. *Int J Med Inf* 1998 Mar;49(1):117-22
- [109] Quantin C, Allaert F-A, Cornet B, Binquet C, Pattisina R, Rodrigues J-M, Gouyon J-B. Comment assurer le chaînage des informations médicales respectant l'anonymat des patients : application à l'étude des pratiques médicales. *J d'économie médicale* 2001, 19, n°7-8, 451-462
- [110] Rainwater JA, Romano PS, Antonius DM. The California Hospital Outcomes Project: how useful is California's report card for quality improvement? *Jt Comm J Qual Improv* 1998 Jan;24(1):31-9
- [111] Ravaud Ph., Giraudeau B., Roux P.M., Durieux P. Les indicateurs de mortalité sont-ils de bons indicateurs de qualité de soins ? *Presse Med* 1999 Oct 2;28(29):1604-9
- [112] Roger VL, Killian J, Henkel M. Coronary disease surveillance in Olmsted County : objectives and methodology. *J Clin Epidemiol* 2002 jun;55(6):593-601

- [113] Romano PS, Rainwater JA, Antonius DM. Grading the graders: how hospitals in California and New York perceive and interpret their report cards. *Med Care* 1999 Mar;37(3):295-305
- [114] Roos LL Jr, Nicol JP, Cageorge SM. Using administrative data for longitudinal research: comparisons with primary data collection. *J Chronic Dis* 1987;40(1):41-9
- [115] Roos NP, Black CD, Frohlich N, Decoster C, Cohen MM, Tataryn DJ, Mustard CA, Toll F, Carriere KC, Burchill CA, et al. A population-based health information system. *Med Care* 1995 Dec;33(12 Suppl):DS13-2
- [116] Roos LL, Nicol JP. A research registry: uses, development, and accuracy. *J Clin Epidemiol* 1999 Jan;52(1):39-47
- [117] Roques F, Nashef SA, Michel P, Gauducheau E, de Vincentiis C, Baudet E, Cortina J, David M, Faichney A, Gabrielle F, Gams E, Harjula A, Jones MT, Pintor PP, Salamon R, Thulin L. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. *Eur J Cardiothorac Surg* 1999 Jun;15(6):816-22; discussion 822-3
- [118] Roques F, Nashef SA, Michel P, Pinna Pintor P, David M, Baudet E. The EuroSCORE Study Group. Does EuroSCORE work in individual European countries? *Eur J Cardiothorac Surg* 2000 Jul;18(1):27-30
- [119] Rosen M, Alfredsson L, Hammar N, Kahan T, Spetz CL, Ysberg AS. Attack rate, mortality and case fatality for acute myocardial infarction in Sweden during 1987-95. Results from the national AMI register in Sweden. *J Intern Med* 2000 Aug;248(2):159-64
- [120] Rosenthal GE. Weak associations between hospital mortality rates for individual diagnoses: implications for profiling hospital quality. *Am J Public Health* 1997 Mar;87(3):429-33
- [121] Rosman DL. The western australian road injury database (1987-1996): ten years of linked police, hospital and death records of road crashes and injuries. *Accid Anal Prev* 2001 Jan;33(1):81-8
- [122] Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. *N Engl J Med* 1996 Jul 25;335(4):251-6
- [123] Seagroatt V, Goldacre M. Measures of early postoperative mortality: beyond hospital fatality rates. *BMJ* 1994 Aug 6;309(6951):361-5; discussion 365-6
- [124] Sear JW, Howell SJ, Sear YM, Yeates D, Goldacre M, Foex P. Intercurrent drug therapy and perioperative cardiovascular mortality in elective and urgent/emergency surgical patients. *Br J Anaesth* 2001 Apr;86(4):506-12
- [125] Semmens JB, Lawrence-Brown MM, Fletcher DR, Rouse IL, Holman CD. The Quality of Surgical Care Project: a model to evaluate surgical outcomes in Western Australia using population-based record linkage. *Aust N Z J Surg* 1998 Jun;68(6):397-403
- [126] Semmens JB, Platell C, Threlfall TJ, Holman CD. A population-based study of the incidence, mortality and outcomes in patients following surgery for colorectal cancer in Western Australia. *Aust N Z J Surg* 2000 Jan;70(1):11-8
- [127] Semmens JB, Wisniewski ZS, Bass AJ, Holman CD, Rouse IL. Trends in repeat prostatectomy after surgery for benign prostate disease: application of record linkage to healthcare outcomes. *BJU Int* 1999 Dec;84(9):972-5

- [128] Sibthorpe B, Kliwer E, Smith L. Record linkage in Australian epidemiological research: health benefits, privacy safeguards and future potential. *Aust J Public Health* 1995 Jun;19(3):250-6
- [129] Silber JH, Rosenbaum PR. A spurious correlation between hospital mortality and complication rates: the importance of severity adjustment. *Med Care* 1997 Oct;35(10 Suppl):OS77-9
- [130] Smith ME, Newcombe HB. Automated follow-up facilities in Canada for monitoring delayed health effects. *Am J Public Health* 1980 Dec;70(12):1261-8
- [131] Smith ME, Newcombe HB. Use of the Canadian Mortality Data Base for epidemiological follow-up. *Can J Public Health* 1982 Jan-Feb;73(1):39-4
- [132] Stenestrand U, Wallentin L. Early revascularisation and 1-year survival in 14-day survivors of acute myocardial infarction: a prospective cohort study. *Lancet* 2002 May 25;359 (9320):1805-11
- [133] Stewart S, Demers C, Murdoch DR, McIntyre K, MacLeod ME, Kendrick S, Capewell S, McMurray JJ. Substantial between-hospital variation in outcome following first emergency admission for heart failure. *Eur Heart J* 2002 Apr;23(8):650-7
- [134] Svartbo B, Olov Bygren L, Gunnarsson T, Steen L, Ribe M. Survival during and after hospitalization: a medical record linkage. *Int J Health Care Qual Assur Inc Leadersh Health Serv* 1999;12 (1):13-7
- [135] The West of Scotland Coronary Prevention Study Group (1995). Computerised record linkage: compared with traditional patient follow-up methods in clinical trials and illustrated in a prospective epidemiological study. *J Clin Epidemiol* 1995 Dec;48(12):1441-5
- [136] Thirion X, Sambuc R, San Marco JL. Anonymity in epidemiological surveys: study and initiation of a new method. *Rev Epidemiol Sante Publique* 1988;36(1):36-42
- [137] Thomas JW, Hofer TP. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. *Med Care* 1999 Jan;37(1):83-92
- [138] Thomas JW, Holloway JJ, Guire KE. Validating risk-adjusted mortality as an indicator for quality of care. *Inquiry* 1993 Spring;30(1):6-22
- [139] Thomas N, Longford NT, Rolph JE. Empirical Bayes methods for estimating hospital-specific mortality rates. *Stat Med* 1994 May 15;13(9):889-903
- [140] Trouessin G. FOIN: a nominative information occultation function. *Stud Health Technol Inform* 1997;43 Pt A:196-200
- [141] Weir N, Dennis MS. Towards a national system for monitoring the quality of hospital-based stroke services. *Stroke* 2001 Jun;32(6):1415-21
- [142] Weir NU, Sandercock PA, Lewis SC, Signorini DF, Warlow CP. Variations between countries in outcome after stroke in the International stroke Trial (IST). *Stroke* 2001 Jun;32(6):1370-7
- [143] Wennberg JE, Freeman JL, Shelton RM, Bubolz TA. Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. *N Engl J Med* 1989 Oct 26;321(17):1168-73
- [144] Yuan Z, Cooper GS, Einstadter D, Cebul RD, Rimm AA. The association between hospital type and mortality and length of stay: a study of 16.9 million hospitalized Medicare beneficiaries. *Med Care* 2000 Feb;38(2):231-45
- [145] McKee M, James P. Using routine data to evaluate quality of care in british hospitals McKee M, James P. *Med Care* 1997 Oct;35(10 Suppl):OS102-11

- [146] Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994 Mar 9;271(10):761-6
- [147] McKee M, Rafferty AM, Aiken L. Measuring hospital performance: are we asking the right questions? *J R Soc Med* 1997 Apr;90(4):187-91
- [148] Epstein AM. Rolling down the runway: the challenges ahead for quality report cards. *JAMA* 1998 Jun 3;279(21):1691-6

Illustrations bibliographiques

A 1 - Ajuster sur les facteurs de risque

Caractéristiques socio-démographiques

Green ^[44] aux USA a comparé les populations de patients des 187 hôpitaux désignés comme "outliers" par la HCFA (Health Care Financing Administration) et celles des 5373 non outliers. Les premiers étaient pour une grande part des personnes âgées de \geq de 85 ans, présentant des diagnostics à haut risque et nécessitant des soins à domicile. Après ajustement, plus de la moitié des outliers ne l'étaient plus.

Sévérité de la maladie

Iezzoni ^[57] démontre l'importance de la prise en compte des critères de sévérité dans le calcul des taux de mortalité. Pour 5 pathologies ou interventions (11 880 infarctus, 7765 pontages, 18 016 pneumonies et 9407 AVC), il a comparé plusieurs techniques de mesure de la sévérité : les taux de mortalité attendue obtenus différaient selon la technique. En outre, la mesure ayant la meilleure capacité à prédire le risque de décès différait selon la pathologie considérée.

Hannan ^[47] a comparé les taux de mortalité intra hospitalière post pontage coronarien obtenus avec des modèles d'ajustement construits à partir de données médico-administratives ou cliniques. La capacité à prédire la mortalité est nettement meilleure avec les données cliniques. L'ajout des 3 seules variables cliniques (fraction d'éjection, ré-intervention, et niveau de sténose du tronc commun) suffit à modifier la hiérarchie entre hôpitaux, certains « bons hôpitaux » devenant « mauvais » et vice versa.

Best ^[5] dans une étude sur les hôpitaux des vétérans non chirurgicaux montre que selon les variables d'ajustement intégrées au modèle, 30 % des hôpitaux changent de place dans le classement effectué sur l'indicateur « mortalité intra-hospitalière ».

Conditions d'admission ou d'intervention

Jarman ^[60] dans une étude sur la mortalité intra-hospitalière portant sur les données de 4 années a constaté que le pourcentage d'admissions en urgence était l'un des facteurs les plus prédictifs du taux de mortalité hospitalière standardisé.

Seagroatt ^[123] montre que les interventions chirurgicales réalisées en urgence ont des profils très différents de celles programmées et considère qu'elles devraient faire l'objet d'une analyse séparée.

Best ^[5] considère également que la prise en compte des conditions d'admission ou d'intervention chirurgicale (programmée ou en urgence) pourrait améliorer le modèle d'ajustement

Un exemple de modèle d'ajustement : EuroSCORE ^{[94] [117] [118]} [site Internet: euroscore.org]

L'EuroSCORE (European System for Cardiac Operative Risk Evaluation) a été développé en Europe entre 1995 et 1999. Les taux de mortalité ajustés sont utilisés comme indicateurs d'efficacité des soins dans le domaine de la chirurgie cardiaque. Les indicateurs développés aux USA ne pouvant pas être utilisés en Europe (en raison des différences entre les patients et les modalités de traitement) un nouvel indicateur a été élaboré par un comité de pilotage, proche de l'association Européenne de Chirurgie Cardiaque. Son objectif est de comparer les résultats des centres, dans le cadre d'une amélioration des performances.

À partir d'une base de données de 19 000 patients ayant subi une chirurgie cardiaque à cœur ouvert dans 8 pays européens (128 centres), une analyse de régression logistique a permis de déterminer les facteurs de risque associés à la mortalité opératoire. Un indicateur ajusté sur les caractéristiques des patients opérés en Europe et standardisé par centre de chirurgie cardiaque a ensuite été mis au point et validé.

17 variables d'ajustement ont été finalement incluses dans le modèle :

Facteurs liés au patient

1. âge
2. sexe
3. BPCO
4. artériopathie périphérique
5. troubles neurologiques
6. antécédents de chirurgie cardiaque
7. créatinine pré-op > 200 µmol/L
8. endocardite
9. état préopératoire critique

Facteurs cardiaques

10. angor instable
11. fraction d'éjection
12. infarctus myocardique récent
13. PAPS élevée

Facteurs chirurgicaux

14. urgence
15. chirurgie cardiaque associée ou non aux coronaires
16. chirurgie de l'aorte thoracique
17. réparation septale post infarctus

(NB : plusieurs de ces variables ne sont recueillies dans le cadre du PMSI).

A 2 - Disposer du nombre d'évènements nécessaire

Mant^[81] a comparé 2 indicateurs de prise en charge de l'infarctus du myocarde : le taux de mortalité et un indicateur de process (% prescription de thrombolyse et de traitements post-infarctus) dans une étude de simulation comparant 2 hôpitaux (sur une base de 450 infarctus du myocarde par service par an, avec une mortalité de 30 % sans traitement et de 14 % en cas de prescription optimale dans 100 % des cas). Il a démontré que sur une période d'un an, il faudrait que la mortalité chute d'au moins 30 % (40 vies sauvées) pour que la différence soit détectable (alors que la mise en évidence des différences de process responsables de cette diminution de mortalité ne nécessiterait le suivi que de 12 patients, soit moins de 2 semaines). Sur une période de 3 ans, il faudrait que la mortalité chute d'au moins 17 % (22 vies sauvées par an) alors que la différence de process correspondante serait détectable par le monitoring de 27 patients par hôpital. L'auteur conclut que même en travaillant sur les données agrégées de 3 années, avec une définition de cas et un ajustement sur la sévérité parfaits, la mortalité spécifique d'une pathologie ne paraît pas être un outil sensible pour comparer la qualité des soins entre hôpitaux (du moins dans les conditions de l'étude et pour la pathologie étudiée). À l'inverse, des audits de courte durée sur les process permettent d'identifier des différences qui ne seraient pas repérées par les taux de mortalité (en outre, les résultats de process sont plus faciles à interpréter et à corriger : par exemple, si on constate que dans un service 30 % seulement des patients reçoivent une thrombolyse dans les 6 heures suivant l'infarctus, l'interprétation est immédiate et la mesure de correction à mettre en place évidente).

Ph. Ravaud ^[111] : Nombre d'interventions pour prothèse de hanche nécessaire pour mettre en évidence une différence statistiquement significative entre les taux de mortalité de 2 hôpitaux					
Effectifs nécessaires par hôpital pour détecter des différences de mortalité entre 2 hôpitaux ($\alpha = 5\%$, $\beta = 20\%$)					
Nombre de patients identiques dans les 2 hôpitaux					
Différence relative Entre les taux de mortalité	Taux de mortalité				
	0,5%	1%	2,5%	5%	10%
50 %	9 383	4 673	1 848	906	435
40 %	15 634	7 785	3 076	1 506	721
30 %	29 525	14 699	5 804	2 838	1 356
20 %	70 321	35 001	13 809	6 745	3 213
10 %	296 839	147 711	58 234	28 408	13 495

B - Qualité des données

Une étude d'évaluation de la qualité de codage dans le cadre du PMSI, réalisée en 1994 dans la région Bretagne (sur 19 hôpitaux et 1814 séjours), a démontré l'hétérogénéité du codage entre les établissements avec un taux moyen d'erreur allant de 7 à 41 % (19 % en moyenne). Les erreurs portaient par ordre de fréquence sur le diagnostic principal (298 séjours), les comorbidités (46 séjours) et les actes (18 séjours) ^[16]. Une autre étude portant sur le codage des séjours en cancérologie d'un hôpital de Rennes montrait que 58 % des séjours changeaient de GHM après recodage, dont 96 % pour une raison médicale (changement de diagnostic principal surtout) ^[76]. Au CHU de Caen, le taux d'erreur est passé de 15,1% en 1995 (65 dossiers sur 431 ont changé de GHM) à 11,2 % en 1997 (70 sur 627 dossiers). Les erreurs portaient sur le diagnostic principal (80 % en médecine, 38% en chirurgie) sur l'oubli d'une pathologie associée classante (20 % en médecine, 33 % en chirurgie) et sur les actes (29 % en chirurgie) ^[22]

Études	D'erreur de codage
Caen 1995	15%
Caen 1997	11%
Bretagne 1994	19% (7 à 41%)
Hospices civils de Lyon	8%
Inter-CHU (rapport 1997)	25% (8 à 44%)

C - Définition de la période d'observation

Jencks ^[62] : le taux de mortalité intra-hospitalière des patients « Medicare » hospitalisés en 1985 était 25 % plus élevé dans l'état de New York qu'en Californie, mais la durée de séjour des patients étant 2 fois plus longue à New York, les taux de mortalité à J30 étaient finalement identiques.

Kahn ^[65] a analysé l'évolution de la mortalité entre 1981-82 et 85-86, dans le cadre d'une étude d'impact de la modification du système de paiement (DRG-based prospective payment system). Ce travail réalisé sur les données de 14 000 patients Medicare montre que l'évolution de la mortalité diffère selon la définition utilisée : diminution de 20 % pour la mortalité intra-hospitalière (de 16,1 % à 12,6 %), de 6 % pour la mortalité à 30 jours (de 16,5 % à 15,4 %), et pratiquement pas de diminution à 180 jours (de 29,6 % à 29 %). Parallèlement, la durée de séjour a diminué de 24%.

Baker ^[3] dans le cadre d'une étude portant sur l'impact de la diffusion de résultats de performance des hôpitaux, a utilisé 3 définitions de la mortalité : intra hospitalière, post-hospitalière survenant dans les 30 jours suivant l'admission et ensemble des décès survenus entre l'admission et J30 (intra ou post hospitalier) : entre 1991 et 1997, les décès se sont décalés de la période intra-hospitalière à la période post- hospitalière : alors que la mortalité intra-hospitalière a baissé de façon significative pour la plupart des pathologies étudiées, la mortalité à 30 jours n'a pas ou peu diminué.

Johnson ^[64] montre que la définition de la mortalité peut modifier le classement des hôpitaux, indépendamment des caractéristiques des patients. Il conclut qu'en l'absence de mesure directe de la qualité des soins, il est difficile de déterminer quelle est la mesure de mortalité la plus informative et quels sont les « vrais » outliers.

Il a analysé les données de 15 288 patients ayant subi un pontage coronarien entre octobre 1993 et mars 1996 dans les 43 hôpitaux Veterans en utilisant 2 définitions :

1 – « mortalité à J30 » (décès quelle que soit la cause, intra ou extra-hospitalier)

2 – « toute mortalité liée à la procédure » (mortalité à J30 + tout décès survenu entre J30 et 6 mois, jugé comme une conséquence directe d'une complication péri opératoire).

En passant de la première à la seconde définition :

- 59 décès supplémentaires s'ajoutent aux 498 décès enregistrés à J30 (soit 557 au total),

- le classement de 86 % des hôpitaux change de 0 à 15 rangs (9 changeant de quartile),

- 5 hôpitaux changent de statut d'outlier, pour 3 d'entre eux sans ajout de décès supplémentaire, par simple effet de reclassement des autres hôpitaux,

- 1 hôpital non outlier devient « high outlier » par la prise en compte de seulement 3 décès supplémentaires survenus à plus de 30 jours.

Garnick ^[33] par contre, comparant la mortalité après pontage coronarien, infarctus du myocarde et insuffisance cardiaque à J30 et à J180 chez des patients Medicare conclut que les 2 mesures ne donnent pas de résultats très différents en termes de classement des hôpitaux (pour les pathologies étudiées).

D - La mortalité est-elle un bon indicateur de la qualité des soins ?

Pour **Krumholz** ^[72], les classements des hôpitaux publiés par « HealthGrades.com » à partir des taux de mortalité à **J30** après infarctus du myocarde parmi les patients Medicare permettent d'identifier des groupes d'hôpitaux qui diffèrent dans la qualité des soins et leurs résultats ; par contre ils discriminent mal les procédures de soins ou les résultats individuels de 2 hôpitaux.

Les données de mortalité après accident vasculaire cérébral publiées par l'Écosse prennent en compte l'âge, le sexe et le statut socio-économique des patients. **Weir** ^[141] a repris les données de 5 hôpitaux couvrant l'éventail d'un classement précédent (étude portant sur 1995-97 – 2 724 AVC vérifiés sur dossier et 1 007 décès à **6 mois**). Il a montré que ce type d'ajustement ne permettait pas d'obtenir des indicateurs de qualité des soins performants. Il a appliqué un second modèle d'ajustement amélioré, déjà validé dans 2 cohortes, prenant en compte l'âge, le fait de vivre seul avant l'AVC, l'autonomie dans les gestes simples de la vie quotidienne, la conservation à l'admission de certaines facultés (peut parler et s'orienter, lever les bras, marcher sans aide). Ce modèle a permis de repérer 1 hôpital avec un taux de mortalité ajusté significativement plus élevé que les 4 autres et dont les procédures de soins étaient effectivement inappropriées. Par contre, il était incapable de différencier des hôpitaux présentant des différences de qualité de soins modérées, quoique importantes cliniquement : malgré des différences notables en terme d'organisation, de spécialisation, de formation (staffs multidisciplinaire), de procédures et de tenue du dossier médical, les 4 hôpitaux avaient presque le même taux de mortalité à 6 mois après ajustement sur le case-mix. Bien que des améliorations soient souhaitables, ces hôpitaux risquent de se satisfaire de ces résultats plutôt que d'y réfléchir de façon critique. Weir conclut que compte tenu des limites de l'utilisation des indicateurs de résultats pour mesurer la qualité des soins, il serait souhaitable d'y associer des mesures évaluant des aspects simples d'organisation et de procédures.

Best ^[5] a réalisé un appariement de patients non chirurgicaux entre 2 établissements (« Veterans Affairs Hospitals »), « outlier » bas et haut (appariement sur le type d'établissement, le diagnostic responsable de la durée séjour, la durée de séjour, la tranche d'âge et la mortalité **intra-hospitalière** attendue). 111 paires, confirmées par une revue du RSS par l'investigateur principal, ont ainsi été constituées. Les dossiers ont été revus à l'aveugle afin d'évaluer la comparabilité des patients d'une même paire, la qualité du codage et la qualité des soins et des procédures. Une relation significative entre la qualité de la prise en charge et le ratio mortalité observée/attendue n'a pas été retrouvée. Les différences observées entre les taux de mortalité reflètent plus des différences entre les patients qu'entre les soins reçus. Pour expliquer cette constatation, l'auteur avance les hypothèses suivantes :

- le modèle de régression logistique utilisé n'est pas performant,
- Il y a beaucoup de facteurs susceptibles d'influer sur le pronostic à court terme qui n'apparaissent pas dans les diagnostics de sortie. L'intégration dans le modèle d'une mesure de sévérité comme l'« apache score » et des conditions d'admission ou d'intervention (programmé ou en urgence) pourrait l'améliorer,
- il existe des problèmes de qualité du codage.

Il conclue en considérant que ce travail doit être poursuivi pour améliorer l'estimation du risque en intégrant dans le modèle d'autres éléments influents et faire en sorte que la qualité des soins reste le principal facteur.

Ce type d'étude pose par ailleurs le problème du « gold standard » de la qualité des soins (les experts évaluant un même dossier ont parfois des interprétations différentes).

Études de simulation

Hofer ^[51] a appliqué un modèle de simulation sur les données de 2 années de 172 hôpitaux de qualité « moyenne » et de 19 hôpitaux de « mauvaise » qualité dans le « Michigan » (5 % versus 25 % de décès évitables). Même après ajustement optimal sur le case-mix, le taux de mortalité intra-hospitalière par DRG ne semble pas une bonne mesure de la qualité. Par exemple, pour l'infarctus du myocarde :

- les taux de mortalité « outliers » élevés ont une valeur prédictive positive de seulement 24 %, les ¾ des hôpitaux concernés ayant en fait une qualité de soins dans la moyenne,
- en regroupant tous les DRG et en situation de parfait ajustement sur le case-mix, cet indicateur présente une sensibilité pour détecter les hôpitaux de mauvaise qualité de 35 % et une valeur prédictive positive de 52 %. L'auteur conclut qu'en dehors de certains diagnostics chirurgicaux, les taux de mortalité hospitalière par DRG ne peuvent pas détecter de façon efficace les mauvais établissements pour la prise en charge des pathologies médicales.

Thomas ^[137] dans une autre étude de simulation en situation « idéale » constate que :

- moins de 12 % des hôpitaux délivrant des soins de mauvaise qualité sont repérés par des taux de mortalité intra-hospitalière anormalement élevés,
- parmi les hôpitaux classés « pors outliers », plus de 62 % sont en fait des hôpitaux délivrant des soins de bonne qualité, soit une sensibilité < 20 % et une mauvaise valeur prédictive (erreur de prédiction > 50 %).

Le pontage coronarien fait exception : 80 % des hôpitaux ayant des taux de mortalité élevés délivrent effectivement des soins de mauvaise qualité et l'erreur de prédiction est de 25 % seulement. Devant ces résultats (non généralisables puisqu'ils dépendent des hypothèses et des pathologies entrées dans le modèle) l'auteur conclut qu'un indicateur de procédures a une meilleure sensibilité pour détecter des différences de qualité et qu'il constitue un meilleur indicateur de performance.

E - Impact de la publication d'un indicateur de mortalité : quelques exemples aux USA

L'expérience de New York : bilan de M. Chassin ^{[18] [19]}

L'état de New York publie annuellement depuis 1989 des taux de mortalité ajustés après pontage coronarien par hôpitaux et par chirurgiens. Parallèlement, il existe une politique de régulation qui limite le nombre d'hôpitaux pouvant effectuer de la chirurgie cardiaque. Chassin tire les leçons de cette expérience et répond aux principales critiques.

A - Impact de la publication :

a) en 14 ans (de 89 à 92), la mortalité a chuté de 41%. Comparé aux autres états, l'État de New York avait en 1992 le taux de mortalité le plus faible et la diminution la plus rapide. La mortalité continuait à être la plus faible en 1999 et était en seconde position en 2000 malgré l'augmentation des patients à risque.

b) il n'y a pas eu d'utilisation des résultats par les compagnies d'assurance ni par les patients.

c) entre 1989 et 1992, 27 chirurgiens ayant une faible activité et de mauvais résultats ont arrêté d'opérer à New York (un des éléments mis à jour par l'analyse des résultats est la relation entre le volume d'activité des chirurgiens et la mortalité).

d) une étude a été réalisée en 2001 dans plusieurs hôpitaux ayant des taux de mortalité supérieurs à la moyenne (outliers) pour voir comment les médecins et les hôpitaux avaient réagi afin d'améliorer leurs résultats. Parmi les mesures adoptées : réorganisation au niveau des services, suspension d'activité de certains chirurgiens, recrutement de chefs de service ou de nouveaux personnels qualifiés, réorganisation du service, programme d'assurance qualité, mise en place d'un département d'anesthésiologie cardiaque, introduction de nouvelles techniques, formation continue etc.

Dans un autre hôpital classé comme « low outlier », la surmortalité était en fait limitée à un sous-groupe de patients très spécifique opérés en urgence et un changement de prise en charge de ces patients a fait chuter la mortalité.

B - Réponses aux principales critiques

a) **les hôpitaux risquent de « sur déclarer » les facteurs de risque** pour améliorer leurs résultats ⇒ On a observé une augmentation de certains facteurs de risque pendant les premières années, conséquence des conseils du comité « advisory » de mieux préciser les définitions. Mais sur 14 facteurs, 9 n'ont pas augmenté. Un audit annuel sur la moitié des hôpitaux vérifie la déclaration des facteurs de risque en retournant aux dossiers ; en cas de discordances importantes un recodage est demandé. L'épreuve de vérité pour savoir si les données sur les facteurs de risque sont pertinentes est de voir comment le modèle de régression logistique utilisé pour l'ajustement fonctionne pour prédire la mortalité. Le modèle s'est avéré extrêmement robuste au cours du temps (C statistics > 0,80 et tests de calibration satisfaisant). Un programme d'audit des données déclarées est essentiel. D'autre part, l'évolution de la prévalence de chaque facteur de risque est surveillée afin de détecter des codage erronés. De toutes façons, les hôpitaux qui déclarent le plus de patients à haut risque ont les taux de mortalité ajustés parmi les plus bas.

b) **les hôpitaux risquent de refuser les patients à haut risque** ⇒ Il y a eu au contraire une diminution du pourcentage de résidents de l'état de New-York se faisant opérer hors de l'État et une augmentation des pourcentages des patients des tranches d'âge à haut risque (65-74 ans et 75-84 ans). Entre 1990 et 1992, le nombre de patients à haut risque a augmenté de 73 %. (NB : d'autres auteurs ne partagent pas cette opinion ^[13]).

c) **les patients risquent de désertier les hôpitaux ayant de mauvais résultats** ⇒ la proportion de patients se faisant opérer dans des hôpitaux à taux ajustés de mortalité supérieurs à la moyenne était de 8,7 % en 1989 et de 9,5 % en 1993 ; la proportion s'adressant à des hôpitaux ayant de meilleurs résultats que la moyenne était de 15,7 % en 1989 et 17 % en 1993.

d) **les données utilisées sont de mauvaise qualité** ⇒ les audits de dossiers effectués n'ont pas montré de différence significative entre la mortalité attendue calculée sur des données auditées et celle calculée sur les données transmises initialement.

d) **la mortalité a diminué également dans d'autres états.** Par exemple Ghali, comparant l'évolution de la mortalité post pontage coronarien dans l'État de New York et dans le Massachusetts, conclue à une diminution comparable (malgré l'absence de diffusion publique des résultats) ⇒ Chassin répond à cet argument en démontrant que les techniques utilisées diffèrent (identification des cas, techniques d'ajustements etc.). Il y a eu une tendance générale à l'amélioration de la survie post pontage entre 1987 et 1992, en moyenne de 5,8%. En 1992 le Massachusetts était

juste au niveau national alors que New York se plaçait nettement au-dessus. Seuls 3 États ayant mis en place une démarche d'amélioration durant la même période avaient des résultats comparables à New York.

Conclusion de Chassin

Au total, le programme a prouvé sa pérennité dans le temps grâce à son intégration en routine, aux publications scientifiques analysant son impact, à l'implication des chirurgiens et des cardiologues, le tout dans un état ayant l'habitude des régulations de son système de soins.

La diffusion de résultats comparatifs galvanise les cliniciens, conduit les administrateurs et les politiques à fournir les ressources supplémentaires nécessaires pour améliorer les services de chirurgie cardiaque.

Mais le système a ses limites :

- déterminer ce qui doit être amélioré entraîne un surcroît de travail. Des indicateurs de résultats ne permettent pas à eux seuls de déterminer les procédures à revoir.
- les améliorations se sont limitées aux hôpitaux pointés comme outliers. Les autres n'ont pas utilisé les données pour s'améliorer et passer de « médiocre » à « excellent ».
- enfin, le programme ne paraît pas généralisable dans l'État de New York sur d'autres sujets ou dans d'autres États (seules quelques expériences similaires ont été menées).

Glossaire

AIREPP	Alimentation informatisée du répertoire des personnes physiques
ALD	Affection longue durée
ANPE	Agence nationale pour l'emploi
ARH	Agence régionale de l'hospitalisation
CAF	Caisse d'allocations familiales
CANAM	La Caisse nationale d'assurance maladie des professions indépendantes
CépiDc	Centre d'épidémiologie sur les causes de décès
CESSI/CNAMTS	Centre d'études des sécurités du système d'information de la Caisse nationale de l'assurance maladie des travailleurs salariés
CIM 10	Classification internationale des maladies, version 10
CISIH	Centres d'information et de soins de l'immunodéficience humaine
CMCD	Certificat médical des causes de décès
CNAF	Caisse nationale d'allocations familiales
CNASEA	Centre national pour l'aménagement des structures des exploitations agricoles
CNAV	Caisse nationale de l'assurance vieillesse
CNAVTS	Caisse nationale d'assurance vieillesse des travailleurs salariés
CNIL	Commission nationale de l'informatique et des libertés
CNRS	Centre national de la recherche scientifique
COTSAM	Conseil pour la transparence des statistiques d'assurance maladie
CREDES	Centre de recherche, d'étude et de documentation en économie de la santé
CREDOC	Centre de recherche pour l'étude et l'observation des conditions de vie
CRI-IFR69	Centre de ressources informatiques de l'institut fédératif de recherche 69
DDASS	Direction départementale des affaires sanitaires et sociales
DHOS	Direction de l'hospitalisation et de l'organisation des soins
DIM	Département d'informatique médicale
DMI2	Dossier médico-économique et épidémiologique de l'immunodéficience humaine
DREES	Direction de la recherche des études de l'évaluation et des statistiques
DRG	Diagnosis Related Groups (groupes apparentés par diagnostic)
EDP	Échantillon démographique permanent
FOIN	Fonction d'occultation d'identifiants nominatifs
GHM	Groupe homogène de malades
GMSIH	Groupement pour la modernisation du système d'information hospitalier
HAD	Hospitalisation à domicile
INSEE	Institut national de la statistique et des études économiques
INSERM	Institut national de la santé et de la recherche médicale
INVS	Institut de veille sanitaire
MCO	Médecine chirurgie obstétrique
MDO	Maladies à déclaration obligatoire
MSA	Mutuelle sociale agricole
NHS	National Health Service
PMSI	Programme de médicalisation des systèmes d'information
RMI	Revenu minimum d'insertion
RNIAM	Répertoire national inter régimes des bénéficiaires de l'assurance maladie
RNIPP	Répertoire national d'identification des personnes physiques
RSA	Résumés de sortie anonymes
SCSSI	Service central de la sécurité des systèmes d'information
SHA	Standard Hash Algorithm
SNIIR-AM	Système national d'information inter régimes de l'Assurance maladie
SSR	Soins de suite et de réadaptation
TAA	Tarifcation à l'activité
URCAM	Unions régionales des caisses d'assurance maladie
URML	Unions régionales de médecins libéraux
URSSAF	Unions de recouvrement des cotisations de sécurité sociale et d'allocations familiales (pas dans rapport intermédiaire)