

**THESE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE**

**PREPAREE A L'U.F.R. Sciences de santé de Dijon**

**Centre d'Épidémiologie et de Santé Publique de Bourgogne (EA4184)**

École doctorale n°554

École doctorale Environnements-Santé

Doctorat de Médecine, santé publique, environnement et société

Par

M. VUAGNAT, Albert Clément

Peut-on utiliser les données du programme de médicalisation du système d'information pour décrire les parcours de soins ?

Thèse présentée et soutenue à Dijon, le 11 septembre 2020

Composition du Jury :

M. GIROUD, Maurice <i>PU-PH, Université de Bourgogne Franche-Comté, EA4184, CHU Dijon-Bourgogne</i>	Président
M. LOMBRAIL, Pierre <i>PU-PH, Université Sorbonne Paris Cité, LEPS EA3412, Hôpital Avicenne, AP-HP</i>	Rapporteur
M. CHATELLIER, Gilles <i>PU-PH, Université Paris Descartes-Sorbonne Paris Cité, Inserm CIC 1418, Hôpital Européen Georges Pompidou, AP-HP</i>	Rapporteur
Mme QUANTIN, Catherine <i>PU-PH, Inserm, UMR 1181 B2PHI, Université de Bourgogne Franche-Comté, CHU Dijon-Bourgogne</i>	Directrice de thèse

## Remerciements

En premier lieu Catherine Quantin, qui m'a fait l'honneur d'accepter que ma contribution à l'activité de recherche du service qu'elle dirige au CHU de Dijon, puisse se convertir en la préparation d'un travail de thèse.

Fabrice Jollant a rejoint le projet et apporté un regard neuf de clinicien qui découvre les systèmes d'informations hospitaliers, avec énergie.

Maurice Giroud et Emmanuel Rusch ont suivi avec bienveillance, dans le cadre du comité, l'avancée des travaux.

Pierre Lombrail et Gilles Chatellier ont accepté d'être rapporteurs du travail rassemblé dans le présent document.

Les responsables de la Direction de la recherche, des études, de l'évaluation et des statistiques, service statistique du ministère des Solidarités et de la Santé m'ont accordé leur confiance tout au long de ces années, Lucile Olier, Christelle Minodier, Nathalie Fourcade, Franck von Lennep, Gwennaëlle Brilhault, Fabien Toutlemonde, Muriel Barlet, Jean-Marc Aubert, confiance sans laquelle les conditions de réalisation de ce travail auraient été très différentes.

Max Bensadon à l'Agence technique de l'information sur l'hospitalisation, m'a toujours accordé le temps nécessaire pour approcher le développement des systèmes d'information, celui du programme de médicalisation du système d'information en particulier.

Les échanges, les savoir-faire, la variété des points de vue, des équipes du CHU de Dijon et de la Direction de la recherche, des études, de l'évaluation et des statistiques, sont une chance et une force.

Ma famille, Sophie, Quentin, Thibault, ont eu la constance, avec un brin de curiosité, de m'accompagner durant ce périple.

*Nec plus ultra*

## Introduction

L'activité de soins médicaux est pour part guidée par des connaissances à valeur générale. Cette médecine savante a été développée progressivement, en Europe à compter de la Renaissance (Grmek et Rey, 1991). Le fil conducteur de cette construction progressive est l'observation des hommes sous l'angle des maladies qui les affectent afin de comprendre de quoi elles résultent, ayant comme présupposé que ce qui les sous-tend est concret et constant au point de pouvoir être exprimé sous une forme mécanistique. Cette compréhension doit en retour permettre d'éclairer, voire d'envisager, les actions thérapeutiques. Les techniques d'observation, initialement sans outil autre que les sens, ont été diversifiées à mesure du développement de savoir-faire en chimie, physique et ingénierie, biochimie, génétique, optique, électricité, ondes, produisant des descriptions toujours plus fines, précises, détaillées du fonctionnement et dysfonctionnement du corps. Toutes ces techniques médicales d'observation peuvent aussi être utilisées pour décrire les effets des thérapeutiques instituées. Ces connaissances factuelles, constituées en corpus scientifique, prennent place aux côtés d'une médecine d'exercice pratique, individuelle, issue du compagnonnage et de l'expérience accumulée, dont la tradition est bien plus ancienne. À défaut d'un équilibre entre la médecine savante et celle de l'individualité de l'action, des échanges de connaissances et points de vue s'opèrent, notamment par des passeurs.

Les observations faites par les médecins lors de leur activité clinique ainsi que les actions thérapeutiques réalisées peuvent être retranscrites par écrit : aide-mémoire pour les consultations échelonnées dans le temps ; communication d'informations relatives aux constatations, aux traitements entrepris, vers d'autres professionnels dans le cadre de la continuité de la prise en charge d'un patient, tant pratique que financière. Ces écrits utilisent le vocabulaire médical du moment, ce qui permet d'être efficace par le biais d'une terminologie synthétique tout en retraçant le lien emprunté vers la représentation des pathologies servant de référence. Ces écrits peuvent, dans des circonstances où du temps dédié leur est alloué, comporter des informations détaillées, par exemple des relevés de mesures, des verbatim ou encore des croquis. L'objet est alors de disposer de documents qui pourront être revus dans un second temps pour préciser ou modifier la description des maladies, leur diagnostic ou leur pronostic, ceci à des fins didactiques, de partage d'expérience ou encore, si les informations d'une série d'observations sont compilées en présupposant un même processus pathologique, d'en déduire des caractéristiques cliniques générales en utilisant des méthodes de représentation, quantification, formalisation mathématiques. Les connaissances cliniques ainsi produites devront être articulées avec le corpus existant, leur intégration le faisant évoluer.

Les données cliniques quantitatives à valeur générale sont devenues une des informations pivot dans la médecine contemporaine. En effet, si l'on admet que la médecine a pour objectif, dans l'idéal, d'apporter par le biais d'une thérapeutique adaptée à la pathologie, une forme de seconde chance à chaque patient qui y a recours (notions de guérison ou encore *restitutio ad integrum*), tout du moins de permettre la reprise au mieux du cours d'une vie personnelle après un événement de santé, des observations en prise directe avec la réalité clinique apparaissent cardinales. Ceci a pour rationnel que la complexité du vivant n'est que très imparfaitement restituée dans les descriptions qui en sont faites, les interactions entre mécanismes, pour ceux connus, étant multiples, sans même envisager ce que la singularité de chaque individu apporte comme incertitude supplémentaire. Malgré cette complexité et ces incertitudes, il y a eu un élargissement des possibilités thérapeutiques, souvent issues d'une approche translationnelle, mais aussi une modération des interventions par le biais de moyens diagnostiques plus performants. Aussi argumentée et attentionnée qu'elle soit, toute nouvelle idée de prise en charge soignante doit être envisagée avec mesure et appréciée en termes de bénéfices et risques dans la réalité clinique. Ce processus de validation concerne aussi les pratiques anciennes, qui

peuvent être remplacées par de nouvelles ou être simplement abandonnées si elles ne modifient pas en mieux le cours des maladies. Il peut aussi être nécessaire de réévaluer des procédures diagnostiques ou des thérapeutiques dans la pratique courante la prise en charge (comprenant le devenir à plus long terme), après des études cliniques initiales concernant souvent une sélection de patients ; le cas échéant les connaissances établies seront complétées, voire amendées. Outre l'aspect temporel de ces connaissances, toujours en mouvement, l'autre difficulté dans la pratique est leur valeur générale, moyenne, relative à un groupe. En effet, être sûr du bénéfice d'une procédure diagnostique ou d'un traitement particulier pour chaque patient considéré individuellement, n'est pas clair en toutes circonstances. Il peut y avoir une zone « grise » pour ce qui n'est pas strictement nécessaire, l'expérience clinique venant alors en aide.

Dès lors, l'organisation de collectes d'observations cliniques structurées apparaît comme un moyen de produire des connaissances cliniques de valeur générale, pouvant contribuer à l'amélioration progressive de la prise en charge des patients. Pour ce faire plusieurs méthodes ont été développées. Dans une première approche, un canevas de recueil est prédéfini car appliquant à une situation précise. Ceci est le cas pour les registres dévolus à l'enregistrement sur un territoire de tous les patients présentant une maladie particulière, l'objet étant de préciser incidence, facteurs de risque, pronostic d'un même processus pathologique et non une collection de curiosités. Un autre exemple est constitué par les cahiers d'observation établis pour une étude clinique, comme le sont les essais tendant à mesurer l'efficacité d'un traitement. Dans le cas où les observations, colligées prospectivement par plusieurs praticiens, doivent être rassemblées en vue d'un résultat plus rapide, un accord entre ces praticiens sur la terminologie employée s'avère devoir être très utile : une classification des maladies, des critères de diagnostic en particulier. Une autre approche consiste en l'indexation des dossiers médicaux en vue de leur revue rétrospective, comme débuté à la Mayo Clinic en 1907, qui couvre alors l'ensemble du spectre des pathologies prises en charge dans l'établissement mais est limité à ce qui a été jugé pertinent de relever lors des soins. Une troisième approche qui étend la notion d'indexation consiste en un recueil systématique et prospectif d'un nombre limité de descripteurs médicaux pour toute prise en charge. L'intérêt de ce recueil continu est de pouvoir y avoir recours à intervalle régulier pour mesurer des évolutions, ainsi que d'être disponible pour des questions incidentes ou documenter des situations cliniques nouvelles.

Un tel recueil d'informations médicales systématique, prospectif, exhaustif a été mis en place dans les établissements de santé en France et est désigné par le vocable « programme de médicalisation du système d'information » (PMSI). Pour chaque séjour, les pathologies et leur prise en charge sont codées sous la forme d'un résumé informatique. Ce recueil a été institué par les services du ministère de la Santé et a comme finalité la valorisation financière de l'activité des établissements de santé. Les données, transmises sous forme déidentifiées par les établissements de santé, sont consolidées dans une base de données nationale, dans un esprit initial assez similaire à la gestion d'archives de comptabilité. Cependant, ces données peuvent être réutilisées avec une finalité autre que la description de l'activité des établissements, dans une optique de recherche ou d'évaluation médicale notamment pour estimer des incidences ou décrire les caractéristiques de pathologies, dans la mesure où celles-ci sont prises en charge en hospitalisation et que la nomenclature des codes ait la précision souhaitée. Divers essais de ce type de réutilisation de la base de données PMSI ont été faits, par exemple en pharmaco-épidémiologie pour dénombrer les effets graves des médicaments (Montastruc et al, 2008), des événements peu fréquents, de forme très diverse, dont il convient de faire le suivi dans le temps et pour lequel un système de signalement a par ailleurs été mis en place, permettant ainsi de conforter voire d'enrichir les estimations réalisées.

À chaque résumé PMSI est associé un couple d'informations dites de chaînage, clé d'identification non signifiante en soi et positionnement temporel relatif, destinées originellement à l'intégration dans le



système d'information national inter-régimes de l'Assurance maladie (SNiiRAM). Ce complément d'information a été utilisé pour conduire des évaluations, comme par exemple l'association temporelle entre la délivrance de Benfluorex et l'hospitalisation pour cardiopathie valvulaire (Weill et al, 2010). Les informations de chaînage peuvent être utilisées par elles-mêmes, sans recours au SNiiRAM : combinées aux résumés du PMSI, elles permettent à priori de reconstituer, de façon anonyme, le parcours de soins hospitaliers des patients, en connectant ensemble les résumés d'un même patient et en restituant leur chronologie. Pour que ceci donne des résultats non parcellaires, il convient que le dispositif de génération des informations de chaînage soit appliqué de façon uniforme et constante dans les établissements de santé. L'originalité du système consiste en une génération strictement locale, sans recours à un service en ligne, et en une consolidation asynchrone de la base de données. À notre connaissance, une évaluation systématique de la cohérence des informations de chaînage figurant dans les fichiers PMSI n'a pas été conduite. Un résultat favorable conforterait alors les études longitudinales destinées à mieux connaître le pronostic des pathologies après leur prise en charge hospitalière, établir des facteurs de risque, apprécier leur évolution dans le temps, ceci à partir d'une source de données alimentée prospectivement, en routine par les équipes hospitalières, publiques et privées.

L'objectif du travail présenté dans ce document est ainsi de mesurer la cohérence des informations de chaînage associées aux résumés PMSI, permettant alors de conforter une réutilisation à des fins de description de parcours de soins hospitaliers. L'hypothèse de travail est que la méthode de génération des informations de chaînage est suffisamment simple et robuste pour être appliquée dans tous les établissements, pour tous les séjours hospitaliers, quelles que soient les circonstances, avec une qualité suffisante pour minimiser la survenue de fragmentation induite des parcours. La seconde hypothèse, lors de l'étude d'une pathologie particulière, est que les nomenclatures utilisées pour renseigner les résumés PMSI comportent des informations d'une précision suffisante et enfin soient complètes. La réutilisation des bases de données PMSI, dont la finalité initiale est le financement des établissements de santé, à des fins de production de connaissances médicales, outre l'efficacité en termes de collecte d'information, offre l'opportunité de disposer de données longitudinales, régulièrement actualisées.

En premier lieu, une revue de l'articulation du système PMSI avec la pratique médicale ainsi que les principes du chaînage des résumés PMSI seront exposés, en prenant appui sur des sources bibliographiques ainsi que des archives administratives. Ces informations ont certainement une utilité dans un contexte d'évolution continue des systèmes d'information automatisés, la présence des échanges internationaux, l'extension des capacités des réseaux, l'irruption des dispositifs électroniques mobiles multimodaux, tous ces aspects ayant un impact sur la pratique médicale.

La comparaison du contenu des fichiers PMSI à des recueils d'informations spécifiques à visée nationale (état-civil et naissances, étude en coupe périnatale, la gestion des transplantations d'organes, les pathologies à déclaration obligatoire) sera présentée dans une seconde section, permettant de mieux cerner les possibilités du recueil PMSI.

Les résultats de vérifications de cohérence générale des informations de chaînage seront présentés dans une troisième section.

Une quatrième section présentera le canevas méthodologique qui a été utilisé pour produire deux études de parcours, l'une sur devenir en termes de réadmission dans les 30 jours qui suivent une intervention chirurgicale, et l'autre sur le devenir des patients dans l'année qui suit une hospitalisation pour tentative de suicide, travaux publiés dans des revues médicales internationales à comité de lecture, et qui clôtureront cette section.

Enfin, les résultats seront discutés et des perspectives esquissées.

## Repères relatifs à la constitution des données médicales hospitalières

### Données du programme de médicalisation du système d'information, PMSI

Sur le modèle du « minimum basic data set » des États-Unis d'Amérique (Linder et Krueger, 1972) et européen (Roger, 1981), un ensemble d'informations est recueilli pour chaque séjour hospitalier dans tout établissement de santé en France. Il s'agit d'un ensemble prédéfini d'informations : démographiques âge, sexe, code postal du lieu de résidence ; relatives au séjour hospitalier numéro d'identification de l'établissement de santé, durée de séjour, mois et année de sortie, mode d'entrée comme de sortie, type d'unité de soins ; médicales diagnostics établis ou traités, actes médicaux réalisés ; informations spécifiques au type de service d'hospitalisation poids de naissance, score de gravité à l'admission en réanimation (IGS2, LeGall et al, 1993), actes de rééducation, hospitalisation sous contrainte, indice d'état général (Karfosky et Burchenal, 1949).

Pour chacune de ces informations, un référentiel des modalités, des règles de codage, et un format d'enregistrement sont définis, rassemblés dans un guide méthodologique de production. Ces informations sont extraites des documents produits par les équipes soignantes lors des étapes de la prise en charge en hospitalisation, au titre de la continuité des soins, et rassemblés dans le dossier médical des patients. Les diagnostics médicaux sont enregistrés sous la forme de codes de la classification internationale des maladies (OMS, 1993), les codes retenus doivent être aussi synthétiques que possible et non reprendre le détail de la présentation clinique. Par comparaison avec ce qui peut se faire lors d'études cliniques, il n'y a pas de transcription de résultats d'examen clinique, biologique, radiologique, de données d'observation et surveillance clinique. La finalité première du recueil d'information étant de décrire l'activité réalisée, seules les pathologies effectivement prises en charge au cours de l'hospitalisation sont retenues. Les actes médicaux sont codés selon une nomenclature spécifiquement française (CNAMTS 2005), la classification internationale des actes médicaux de 1978 (1980 pour la version française), n'ayant pas eu de suite, donc sans les nombreuses évolutions des techniques ayant eu lieu depuis. Les actes médicaux opératoires effectués doivent figurer, car pris en compte dans la valorisation des séjours via l'algorithme de groupage (cf. infra), l'enregistrement des autres actes médicaux est à la discrétion des établissements de santé, ils le sont alors dans une visée documentaire. Les codes des actes médicaux sont définis comme décrivant l'ensemble d'une procédure thérapeutique ou diagnostique et non pas le détail d'actions élémentaires, ceci dans une optique de facturation traçable, un tarif de responsabilité défini pour chacun d'eux. Ne figurent pas dans les données PMSI, les actes infirmiers ou des autres personnels soignants, ni les médicaments et perfusions, à l'exception de certaines chimiothérapies anti-cancéreuses de caractère onéreux. Les nomenclatures des diagnostics comme celle des actes médicaux sont destinés à des spécialistes, seuls les libellés sont donnés, pas de description détaillée clinique ou opératoire, à l'exception toutefois d'une présentation spécifique du chapitre des pathologies mentales (OMS, 1994)

Ce descriptif condensé, structuré et normalisé de chaque séjour est désigné sous le terme de « résumé PMSI ». Le recueil est organisé réglementairement par le ministère de la santé. La finalité première est la valorisation financière de l'activité de soins réalisée par les établissements de santé. Cette valorisation se fait par l'application aux informations contenues dans chaque résumé PMSI, d'un algorithme publié dont le résultat est l'affectation dans un groupe d'une classification qui couvre l'ensemble de l'activité des soins en hospitalisation ; l'application de l'algorithme se fait au moyen d'un logiciel, les résumés PMSI étant sous la forme d'enregistrements informatiques. Pour chaque groupe, un tarif est défini. Le tarif est déterminé en fonction d'une échelle de coûts observés, actualisée chaque année, qui donne une valeur relative, et de telle sorte que la somme des dépenses en soins d'hospitalisation ne dépasse pas le montant fixé par le parlement pour chaque année civile. Ce système de valorisation d'activité par

un ensemble de forfaits constitue une forte incitation à ce que chaque établissement constitue un résumé PMSI pour tous les séjours réalisés mais aussi à ce que chaque résumé comporte toutes les informations correspondant à chaque séjour, i.e. des informations exhaustives et complètes. L'Assurance maladie, qui supporte la valorisation financière des séjours, effectue en se rendant dans les établissements de santé, des contrôles du contenu d'une sélection de résumés PMSI, en confrontant ce contenu aux éléments des dossiers médicaux correspondant. Le cas échéant, un trop perçu est réclamé. Cette disposition régule un éventuel sur codage des résumés PMSI. Un second dispositif de régulation par l'Assurance maladie est constitué par la définition de critères cliniques minimaux nécessaires (instabilité de l'état, actes nécessitant un plateau technique hospitalier, dépendance) pour que la prise en charge puisse être considérée comme relevant de l'hospitalisation et non de soins de consultation externe. Les contrôles portent également sur la vérification de l'application de ces critères de facturation. Ce dispositif a pour effet de normaliser le périmètre de l'activité décrite par les résumés PMSI.

Chaque établissement de santé constitue au fur et à mesure les résumés PMSI correspondant à l'activité réalisée, et les transmet par voie électronique à l'Agence technique de l'information sur l'hospitalisation. Celle-ci consolide l'ensemble des résumés PMSI reçus en base de données nationale, pour le compte du ministère de la Santé, responsable de ce traitement informatique de données personnelles. Cette base de données permet de décrire et suivre l'activité d'hospitalisation, sa valorisation financière, de préparer les évolutions tarifaires, de les adapter aux évolutions des profils de prise en charge. Bien que ne comportant pas d'information nominative, les informations spatiales et temporelles figurant dans les résumés PMSI (numéro d'établissement, mois de sortie, âge, code postal, ...) sont en nombre et en précision suffisants pour permettre l'identification par recoupement d'une personne particulière (Cnil 2000) et dès lors faire le lien avec les informations sur sa santé (diagnostic et actes médicaux des résumés PMSI correspondants), informations de caractère privé et protégées par la loi, avec en particulier des dispositions relatives au secret professionnel. En conséquence, l'accès et l'utilisation de la base de données PMSI, bien que de caractère administratif, sont règlementés : la finalité doit être la réalisation d'études et des évaluations d'intérêt général, par du personnel habilité au sein de structures autorisées ; la sécurité informatique doit être garantie et, plus récemment, la traçabilité des actions a été instituée. Les travaux présentés dans le présent document prennent appui sur la base de données PMSI, et ont été réalisés dans ce cadre réglementaire. Ce cadre est suffisamment large pour permettre de produire, au-delà du pilotage financier, des descriptions statistiques des recours aux soins hospitaliers par pathologie, un complément utile à la connaissance de l'état de santé de la population. Pour qu'une utilisation de cet ordre de la base de données PMSI soit contributive, le condensé d'information que constituent les résumés PMSI doit être suffisant pour la description clinique souhaitée, et d'autre part, la place de la prise en charge en hospitalisation doit pouvoir être appréciée. Si ces réserves ne sont pas bloquantes, les travaux de recherche et d'évaluation peuvent utilement prendre appui sur la base de données PMSI eu égard à ses caractéristiques : l'exhaustivité du recueil, imposé par le contexte donné supra, évite les situations de censure, qu'il n'est pas toujours aisé de qualifier de non informatives, et la normalisation du codage des informations à travers des référentiels de taille prédéterminée, permet une mise en œuvre informatique déterministe, vérifiable et réutilisable. Pour cerner au mieux la portée du contenu de la base de données PMSI, des éléments d'articulation entre ce système d'information et son contexte, peuvent apporter l'éclairage souhaité.

La définition du contenu des résumés PMSI, comme de ses prédécesseurs « minimum basic data sets », répond à l'objectif de pouvoir être appliqué à tout type d'hospitalisation, et de comporter une information sur la pathologie prise en charge. Un point d'équilibre doit donc être trouvé entre complexité, détail des informations d'une part et une simple indicatrice d'admission en hospitalisation d'autre part, à l'instar de la définition de seuils distinguant des valeurs normales de pathologiques, faite dans ce cas pour être efficace. En effet, il s'agit de mettre en place une routine à travers de nombreuses

équipes, diverses, ce avec de l'idéal un impact sur le fonctionnement quotidien aussi réduit que possible. S'il y a déjà des compréhensions communes, des méthodes de recueil existantes, du savoir-faire préexistants, il peut être pragmatique de les prendre en compte dans l'organisation du recueil, d'autant plus accessible qu'il y a moins de nouveautés à assimiler. De fait, le recueil PMSI n'évolue que très progressivement. Pour les services de soins, l'intérêt du codage de l'activité peut être porté par le fait de disposer en retour d'une description des pathologies prises en charge mais aussi, étant entendu que la forme synthétique retenue pour les résumés PMSI ne permet pas l'étendue des descriptions pertinentes pour la clinique, de constituer une clé d'indexation des dossiers médicaux, qu'il sera alors possible de revoir dans un second temps pour parfaire la description d'une situation clinique particulière, en vue d'un travail de recherche ou d'évaluation clinique. Pour ce dernier type d'utilisation, une valeur prédictive positive élevée des informations des résumés PMSI est souhaitable, ce qui engage les services à faire un codage précis et complet des séjours. Pour aider à la mise en place de cette nouvelle organisation, la fonction de médecin responsable de l'information médicale a été créée, avec initialement un financement spécifique révélateur de l'engagement du ministère de la santé pour installer le recueil des résumés PMSI au sein de chaque établissement de santé. Ce médecin a pour rôle de coordonner la collecte par des échanges avec les autres médecins, notamment autour du codage des informations et d'éléments de documentation des dossiers médicaux. Il peut aider à définir les formats utiles de retour d'information et de bilans d'activité médicale. Dans la pratique, sa contribution peut parfois même être aussi technique que la participation au développement du système d'information métier de l'établissement. Il est chargé de la transmission des résumés PMSI à l'Agence technique de l'information sur l'hospitalisation au moyen de logiciels à installer sur un équipement micro-informatique (qui fait son apparition, à la fin des années 1980, dans les hôpitaux, où étaient installés des gros systèmes de gestion (Cour, 1992) comme le système Fontenoy) et de veiller à la cohérence du contenu des résumés PMSI. Le droit afférant au secret professionnel et médical en particulier, a été adapté afin de lui permettre d'exercer utilement sa fonction en lui autorisant de revoir des dossiers médicaux de patients pour lesquels il n'a pas concouru aux soins (Ordre, 2000). Le ministère de la santé, par l'intermédiaire de l'Agence technique de l'information sur l'hospitalisation, compile dans un guide mis à jour régulièrement, les instructions relatives au renseignement des résumés PMSI. Avec cette organisation générale, il est escompté un recueil plus robuste et fluide.

Si la généralisation d'un relevé systématique et normalisé d'information pour chaque hospitalisation a constitué une nouveauté, le principe d'un recueil prospectif de données médicales structurées et en routine clinique, en vue de faire des constats quantifiés permettant des actions et non pas seulement d'entériner la gestion (norme de fonctionnement, décompte des admissions), avait déjà été appliqué en France en diverses circonstances. Ceci a pu concerner un type de pathologie, les maladies infectieuses à déclaration obligatoire (1892), les causes médicales de décès (1937), des registres notamment de cancers (1975), mais aussi certaines spécialités médicales par exemple la réanimation, l'obstétrique (Robelet, 1999) qui ont en commun des décisions thérapeutiques vitales parfois dans l'urgence, ou encore des services, voire des établissements entiers (Valois, 1970), souhaitant pouvoir revoir des dossiers sur des critères médicaux. En 1942, fut organisé dans les hôpitaux généraux publics de Nancy et sa région, un recueil très similaire à celui des résumés PMSI : pour tous les patients sortants un bulletin statistique individuel était renseigné, permettant alors d'établir des statistiques d'activité en termes de pathologies prises en charge. Ceci fut ensuite étendu par l'Insee pour les années 1945 et 1946 dans plusieurs régions (Chaperon et Trehony, 1980). Ce système n'a pas perduré, malgré des tentatives en 1958 et 1968-1974, et a été remplacé, avant que le recueil PMSI ne soit institué, par des enquêtes transversales de morbidité hospitalière (1985, 1992), organisées par le service statistique du ministère de la santé, qui fournissent un ensemble de descriptions médicalisées et que leur répétition permet de traduire en tendances évolutives, ce qui peut donner des éléments quantitatifs pour la planification sanitaire. Des enquêtes transversales sur une pathologie ou un type de recours, sont aussi organisées :

enquête nationale périnatale (1972), sur les urgences hospitalières (2002). Ceci ne signifie pas pour autant qu'une pratique généralisée de relevé d'informations médicales à visée d'évaluation ou de recherche préexistait et sur laquelle le recueil PMSI aurait pu prendre appui. Par ailleurs, en terme de méthode, nombre des recueils mentionnés supra se placent, dans la lignée des transcriptions administratives, i.e. faire correspondre la situation qui se présente, au mieux des informations disponibles sur le moment, à l'une des modalités prédéfinies afin de ne pas être bloqué dans la poursuite de son travail de son action. La technique de recueil est celle d'une transcription : les informations sont prises dans des documents formalisés ou doivent leur correspondre s'ils produits simultanément (par exemple sélection d'un code d'acte médical pour initier un compte-rendu d'intervention) ; il n'y a pas de recueil de données cliniques source (mesures, images, tracés, biologie, ...) qui ouvrent sur la possibilité de réinterprétation ; par contre, des informations formalisées, notamment pour ce qui concerne les diagnostics et les actes médicaux, emportent avec elles une part de motif de recours à l'hospitalisation ou d'indication thérapeutique, l'action des équipes de soins. Pour mener à bien cette tâche de transcription, il n'est pas strictement nécessaire de disposer d'une vue générale du contexte du recueil d'information. Cependant, la transcription peut parfois nécessiter une interprétation, le résultat sera alors fonction de la compréhension, du savoir-faire de celui réalise le recueil. Pour les données de démographie, comme de séjour des résumés PMSI, les référentiels sont pour la plupart de compréhension commune et disponibles (état-civil, calendrier, localisations, ...). Par contre, pour le codage des diagnostics comme celui des actes, celui-ci peut, même pour des cliniciens spécialistes les plus expérimentés, prêter à un choix au mieux. De fait, il est admis, même pour les mesures quantitatives les plus élémentaires, une imprécision au-delà même de la variabilité en soi des situations (Laferriere et al, 1978). En conséquence, pour un usage de la base de données PMSI à des fins d'évaluation ou de recherche, il convient d'apprécier l'homogénéité des situations recouvertes par chaque information, tant lors du projet, que de l'analyse des données, ou encore de l'écriture de la synthèse des résultats. Les données PMSI, du fait de leur existence et de leurs caractéristiques, sont progressivement mises à contribution au-delà d'être un vecteur de financement des établissements de santé. Sont ainsi alimentés, des descriptions des recours au système de soins, de l'état de santé de la population, mais aussi des indicateurs tendant à approcher la qualité des soins (évolution des pratiques, des techniques). La diffusion de ces résultats peut parfois se faire vers le grand public, ce qui ajoute à l'inconfort des cliniciens vis-à-vis du recueil PMSI, corset financier, jugement sur les soins réalisés, et non une information commune, utile pour la pratique. Des études conduites par des équipes pluridisciplinaires sont cependant réalisées.

Le chaînage des résumés PMSI successifs d'un même patient, sans afficher son identité, ouvre des possibilités d'études longitudinales, sur le devenir des patients pris en charge en hospitalisation, les facteurs de risque associés, parfois des arguments pour l'évolution des pratiques ou encore pour informer le public.

### **Chaînage des résumés PMSI**

Les techniques mécanographiques, a fortiori informatiques, permettent de rapprocher des enregistrements d'informations recueillies auprès d'une même personne, au moyen de méthodes automatisées qui prennent appui sur des traits d'identification (Newcombe et al, 1959). Ces rapprochements peuvent servir plusieurs objets : dans des bases de données sans identifiant directement signifiant pouvoir rassembler les enregistrements par personne pour un suivi longitudinal ; identifier des enregistrements en double et se prémunir du biais induite par des doubles comptes ; enrichir les informations d'une base de données avec celles rassemblées dans un autre contexte en évitant la duplication de collectes et en étant ainsi parcimonieux des ressources ; consolider les informations depuis plusieurs sources pour des données plus robustes. Dans la pratique courante, la

mise à jour d'information se fait après une recherche individuelle d'un dossier, d'un enregistrement. Pour servir de traits d'identification à visée de chaînage, les informations doivent être propres à la personne, être stables dans le temps, avoir une définition commune, être simple d'utilisation. Les traits d'identification sont de différentes natures, directement identifiantes dans la vie courante comme les informations d'état-civil ou qui en sont déduites comme le numéro d'inscription au répertoire national des personnes physiques dit « numéro de sécurité sociale » (Carmille, 1941), ou encore une adresse postale, électronique, un numéro de téléphone, ou plus indirectement encore une combinaison unique d'informations spatiales et temporelles, par exemple date, heure et lieu d'admission en hospitalisation (Lawson et al, 2013). Il est encore distingué des méthodes de chaînage dites déterministe, ne sont rassemblés que les enregistrements avec une correspondance exacte des informations, de méthodes dites probabiliste, appliquant une pondération aux éléments d'un profil d'identification afin de tenir compte d'une non constance dans leur enregistrement et diminuer en conséquence les ruptures indues de chaînage, les pondérations étant issues du contexte de collecte des informations ; les méthodes déterministes utilisent généralement des informations prédestinées au chaînage, comme par exemple un numéro d'identification ; les méthodes probabilistes peuvent prendre en compte une combinaison unique d'information, par exemple spatiales et temporelle. Enfin, si la mise en œuvre d'un rapprochement automatisé d'informations apporte aux destinataires avec efficacité des informations utiles pour établir des connaissances nouvelles, ceci doit être fait pour une finalité licite, d'intérêt général, dans le respect du droit des personnes, des malades en particulier. En France, l'utilisation de l'informatique pour le traitement des données personnelles est encadrée depuis 1978 par la loi ; la question est celle du droit à la vie privée, individuelle et de la non-discrimination (Scheuren, 1999).

Le chaînage des résumés PMSI se fait au moyen d'un couple d'informations non significatives en soi, un pseudonyme et un délai. Le pseudonyme est le résultat de l'application d'une fonction de hachage au triplet d'informations numéro de sécurité sociale de l'assuré, date de naissance et code sexe du patient. Les fonctions de hachage effectuent une séquence de traitement binaire de l'information en entrée, qui la recombine en un résultat irréversible. Elles sont élaborées de façon pragmatique en combinant diverses opérations binaires, translations, permutations, additions, soustractions afin de donner un résultat non inversible. La séquence de traitement est complètement déterminée et n'utilise que les informations en entrée, ainsi le même triplet d'information donnera toujours le même pseudonyme. Les fonctions de hachage considèrent toute l'information binaire en entrée et la redistribuent (comme le ferait un pétrin mécanique ou encore lorsque l'on fouette une sauce béarnaise), une modification même ponctuelle des données en entrée donnera un résultat différent en sortie. Pour des informations courtes en entrée, c'est-à-dire d'une longueur similaire au résultat de la fonction de hachage (il n'y a pas de compression d'information) comme dans le cas présent, il est admis qu'il n'y a pas de collision (i.e. des informations différentes donnant le même hachage en sortie). Ainsi, par cette méthode, les traits d'identification sont remplacés par une chaîne de caractères non significative mais constante et unique pour le même triplet d'informations source. Le triplet d'informations source étant constant pour un même patient car correspondant à son état civil, le pseudonyme qui en est calculé aussi ; il convient toutefois de noter l'existence de changements de statut vis-à-vis de l'Assurance maladie pour les enfants et les conjoints. La réglementation qui encadre l'utilisation du numéro d'inscription au répertoire des personnes physiques, permet son utilisation pour la facturation et le recouvrement des dépenses de soins, donc aussi dans les établissements de santé. Ce numéro ne peut figurer, sauf autorisation spécifique de la loi ou par la Cnil, dans des fichiers informatiques à côté de données de santé. Les résumés PMSI, pour inclure le moins possible d'informations indirectement identifiantes, ne comportent pas de date, seulement le mois et l'année de sortie ainsi que la durée de séjour. Pour permettre cependant un positionnement chronologique relatif des résumés PMSI d'un même pseudonyme, le délai en nombre de jours entre une date fictive calculée à partir du pseudonyme et la date d'admission en hospitalisation, constitue la seconde information de chaînage des résumés PMSI.

Les informations de chaînage sont générées par un logiciel développé par l'Agence technique de l'information sur l'hospitalisation et diffusé aux établissements qui doivent le mettre en œuvre pour transmettre les résumés PMSI dans le cadre de la valorisation de leur activité (Ministère de l'emploi et de la solidarité, 2001). Ce logiciel comprend un module développé par le Centre d'études des sécurités du système d'information de la Caisse nationale de l'Assurance maladie, qui génère le pseudonyme. Le logiciel vérifie le format de chaque élément du triplet d'informations avant l'application de la fonction de hachage et le cas échéant signale une non-conformité à travers une série d'indicateurs désignées par le vocable « codes retour ». Leur description est présentée dans le tableau 1. En effet, sans ces indicateurs, il n'est pas possible de distinguer dans leur forme des pseudonymes générés à partir d'informations avec des erreurs de format ; ces pseudonymes sont alors non utilisables à des fins de description statistique. Le dispositif est appliqué à tous les résumés PMSI depuis 2001, ce qui permet de disposer d'une base de données de résumés PMSI exhaustive et chaînée sur un mode déterministe, et dans la mesure où la procédure de constitution du triplet d'informations est constante et conforme dans tous les établissements, la possibilité technique de décrire, notamment avec les codes de pathologies et des actes médicaux, le parcours de soins hospitaliers.

La version initiale du chaînage des résumés PMSI a été développée en 1996 pour accompagner, à la demande de la Caisse nationale de l'Assurance maladie, le financement de l'obligation de recueil PMSI faite aux établissements de santé privés à but lucratif. Pour éviter de faire figurer le numéro d'assuré social dans les résumés PMSI, le Centre d'études des sécurités du système d'information (CESSI/CNAMTS) a conçu et fourni dès 1996 une fonction d'anonymisation appelé FOIN (Fonction d'Occultation d'Information Nominatives) pour la mise en place du PMSI dans les établissements privés, sur recommandation de la CNIL (qui a suggéré l'utilisation de l'algorithme développé par le Pr Catherine Quantin, (Dusserre, Quantin, 1995; Quantin et al, 1998a; Quantin et al, 1998b; Quantin et al, 2005; Quantin et al, 2014a), sur demande du MES/DH8 et après expertise du SCSSI. Les informations de chaînage, appliquées à ces seuls établissements, devaient permettre la mise en correspondance des résumés PMSI avec le résumé de facturation de séjour correspondant. De la consultation, expressément autorisée, des archives du ministère de la santé de cette période (mission des archives de France), il ressort que le chaînage des résumés PMSI devait aussi permettre leur connexion, sans recours au numéro de sécurité sociale, aux bases de données nationales déidentifiées de l'Assurance maladie et dès lors la possibilité de faire un suivi des dépenses par groupe de pathologies dans le cadre de la gestion du risque. Pour la génération du pseudonyme, les sécurités suivantes sont appliquées : le hachage est réalisé par la fonction SHA-1 (U.S. Department of commerce, 1995) la plus robuste disponible, l'ajout aux informations en entrée d'une chaîne de caractères inconnue aux utilisateurs pour éviter une attaque par dictionnaire, une troncature du résultat en sortie de la fonction de hachage, la réalisation d'un second hachage selon le même procédé avec une autre chaîne inconnue de caractère au moment de la réception à l'Assurance maladie des lots de résumés PMSI télétransmis (télétransmission sécurisée par la procédure xmodem : authentification point à point, vérification intégrité, journalisation), ni les pseudonymes ni l'information de positionnement chronologique relatif ne figurent dans la copie des fichiers télétransmis effectuée sur le poste informatique. Les chaînes inconnues sont reconstruites à partir de deux fragments, l'un généré par le ministère de la Santé, l'autre par la Caisse nationale d'Assurance maladie ; l'intégrité et l'authentification des fragments est vérifiée automatiquement par le module FOIN au moment de la reconstruction de la chaîne inconnue. Pour la fonction SHA, il est pris appui sur un programme écrit en C mis à disposition librement sur un forum et conforme aux spécifications techniques (Gutmann, sci.crypt, 1992) ; en outre, ce programme teste l'ordre de chargement en mémoire des octets de chaînes de caractère (« endianness ») fonction du système et le cas échéant, réalise la permutation pour que l'application de l'algorithme SHA-1 soit normalisée. La fonction SHA-1 considère l'information en entrée par blocs de 512 bits, soit 64 octets ou caractères codés en Ascii. Dans le cas présent un seul bloc est suffisant. Il y a donc peu de compression

selon le schéma de Merkle-Damgard, diminuant d'autant le risque de collision (Bouillaguet, 2016). Le résultat du hachage est transposé par blocs de 5 bits en un alphabet de 32 caractères, chiffres et lettres majuscules. Les pseudonymes ont ainsi une longueur de 16 caractères sans caractère spécial. Les fonctions de vérification, reconstruction, hachage, transposition, sont intégrées dans le même module, FOIN (pour fonction d'occultation des informations nominatives) développé le Centre d'études des sécurités du système d'information de la Caisse nationale de l'Assurance maladie. Ces éléments de sécurité ont obtenu l'aval du service central de sécurité des systèmes d'information (SCSSI), nécessaire en ces années y compris pour l'utilisation de fonctions de cryptologie ne réalisant que la signature (SHA-1) ou testant authenticité et intégrité, ainsi que celui de la Commission nationale de l'informatique et des libertés (Cnil) sur le principe de ce chaînage par pseudonyme irréversible, ce qui est suffisant, la finalité n'étant pas un retour aux dossiers médicaux mais la réalisation de descriptions statistiques.

La télétransmission des résumés PMSI à la Caisse nationale de l'Assurance maladie a été remplacée en 2001 par une transmission à l'Agence technique de l'information sur l'hospitalisation en charge de la consolidation de la base de données PMSI nationale. À compter cette année, le chaînage est appliqué à l'activité de tous les établissements de santé, publics comme privés. Le module FOIN est adapté, le code sexe remplaçant l'information rang de bénéficiaire dans le triplet d'informations en entrée de la fonction. En effet, dans les établissements publics de santé, alors financés par dotation globale, l'information de facturation à l'Assurance maladie, rang de bénéficiaire, n'y était pas disponible. Cette modification dans le triplet d'information en entrée a pour conséquence cependant de ne pouvoir distinguer les pseudonymes des enfants jumeaux et de même sexe. Pour documenter la chaîne de caractère inconnue utilisée lors de la génération des pseudonymes, une indicatrice la caractérisant est ajoutée à la suite du pseudonyme (« C » pour le premier hachage, « D » pour le second). Le système de sécurité apporté par la fonction FOIN, est complété par un réordonnement aléatoire de l'ordre des résumés PMSI au moment de la réception pour empêcher la constitution d'une table de correspondance entre pseudonymes transmis par les établissements de santé et ceux issus de la deuxième application du module. Il est à noter que le module de génération des pseudonymes ne prévoit pas au moment de l'utilisation, ou tout du moins après l'installation sur un ordinateur, la réalisation systématique de la génération d'un pseudonyme test à partir d'un triplet d'informations de test afin de vérifier que le résultat obtenu est bien celui attendu, comme cela se fait par exemple pour le calibrage des automates de laboratoires. Le risque de génération de pseudonymes différents pour des triplets identiques (perte induite de chaînage) conséquence de la diversité potentielle des environnements informatiques, dans les établissements de santé en particulier, a tenté d'être circonscrit : le module est intégré dans un logiciel qui lit les données du triplet en tant que texte ; le logiciel ne peut fonctionner que sur des ordinateurs utilisant un système d'exploitation précis ; les informations en entrée sont constituées exclusivement de chiffres qui, lorsqu'ils sont codés comme du texte, correspondent aux codes binaires 0x30 à 0x39 respectivement pour 0 à 9, ces valeurs sont constantes entre différentes convention de codage binaire des caractères : Ascii 7bits, les extensions Ansi 8bits (Latin-1/Iso-8859, Windows-1252 pour les plus fréquentes), UTF-8. L'utilisation de l'algorithme SHA-1 pour la génération des pseudonymes, date de 1996. D'un point de vue cryptographique, cet algorithme est considéré depuis quelques années comme non sûr, en particulier pour la signature de certificats de sécurité informatique, certificats ensuite utilisés pour le chiffrement et la sécurisation des échanges via les réseaux informatiques. En effet, des méthodes plus efficaces que la force brute (qui est de parcourir systématiquement toutes les valeurs possibles), ont été proposées afin de produire un résultat de hachage identique alors que les informations en entrée sont différentes. Dès lors la fonction signature de l'algorithme SHA-1 n'est plus absolument sûre. Cependant, pour l'usage qui en est fait pour le chaînage des résumés PMSI, i.e. la génération stable d'un pseudonyme, ceci n'est pas remis en cause (Gutmann et al, 2005, Preneel, 2010). Enfin, les pseudonymes issus de la succession de hachage avec le module FOIN doivent aussi correspondre à ce qui figure dans le système national inter-régimes de l'Assurance maladie (SNiiRAM),



pour rendre possible la description des parcours de soins hospitaliers élargis aux soins réalisés en consultation. Le système SNIIRAM comporte aussi le statut vital, information communiquée par l'état civil dans le cadre de l'aide à la gestion des droits par l'Assurance maladie. Lors du développement d'un projet d'étude de la mortalité post-hospitalière (Vallet et al, 2004, ministère de la santé, 2004) il a été constaté que pour correspondre aux pseudonymes figurant dans le SNIIRAM, l'ordre des éléments de la date de naissance (jour, mois, année) dans le triplet d'informations en entrée devaient être permutés pour les données PMSI, sans quoi aucune correspondance ne se faisait ; le logiciel comportant le module foïn a été modifié en conséquence et a été appliqué à toutes les transmissions de résumés PMSI depuis 2004.

Ainsi, le dispositif technique de génération des pseudonymes destinés au chaînage des résumés PMSI est stable depuis 2004, avec des garanties de mise en œuvre uniforme dans les établissements de santé, diminuant le risque de divergence de génération des pseudonymes ayant pour conséquence la rupture induite du chaînage des résumés d'un même parcours hospitalier d'un patient. Les caractéristiques de ce dispositif permettent dès lors d'envisager la constitution asynchrone d'une base de données de résumés PMSI qui comportant des informations de chaînage traduisant chaque parcours de soins hospitaliers. La stabilité des pseudonymes générés est encore accrue par l'intégration de plus en plus routinière dans les processus de gestion administrative des soins, du recours à la carte à puce propre à chaque assuré social (objet d'identification et d'information sur les droits et non média d'échanges Brams, 1990), ce système contribue aussi à la qualité et la constance des informations en entrée de la fonction de hachage. Le mode de calcul de l'information de positionnement chronologique relatif à une date fictive non connue de l'utilisateur, a été modifié en 2006 : la référence à l'année en cours, dont la prise en compte pouvait être ambiguë notamment en début et fin d'année civile, étant alors remplacée par une année fixe dans une plage définie, donnant un résultat plus simple d'emploi.

Tableau1 : liste des codes retour associés à la génération des informations de chaînage

Libellé	Année		2008	2009	2010-2012	2013-	
	Champs PMSI	2001-2005	2006-2007	MCO, HAD*	SSR, RIM-P*	Tous	Tous
	Code	Tous	Tous	Code	Code	Code	Code
<b>1er code retour : contrôle "n° sécurité sociale"</b>							
Pas de problème à signaler	0	0	0	0	0	0	0
Format incorrect			1	1	1	1	
Longueur incorrecte (différente de 13)	1	1					
Renseigné avec la valeur manquante (i.e. XXXXXXXXXXXXX)	2	2	2	2	2	2	
Présence de caractères non numériques	3	3					
Format incorrect (i.e. ne commençant pas par 1 ou 2)	4						
Format incorrect (i.e. ne commençant pas par 1, 2, 5, 6, 7, 8)		4					
N° non conforme suite au contrôle de la clé associée au n° d'assuré social			3	3	3	3	
Traitement impossible (i.e. absence de génération d'une clé de chaînage)	9	9	9	9	9	9	
<b>2ème code retour : contrôle "date de naissance"</b>							
Pas de problème à signaler	0	0	0	0	0	0	0
Longueur incorrecte (différente de 8)	1	1					
Renseignée avec la valeur manquante (i.e. XXXXXXXX)	2	2	2	2	2	2	
Présence de caractères non numériques	3	3					
Date incorrecte (i.e. ne peut être interprétée en tant que date au format jj/mm/aaaa)	4	4	1	1	1	1	
Traitement impossible (i.e. absence de génération d'une clé de chaînage)	9	9	9	9	9	9	
<b>3ème code : retour contrôle "sexe"</b>							
Pas de problème à signaler	0	0	0	0	0	0	0
Renseigné avec la valeur manquante (i.e. X)	1	1	2	2	2	2	
Valeur illicite (i.e. n'appartenant pas à {1, 2})	2	2	1	1	1	1	
Traitement impossible (i.e. absence de génération d'une clé de chaînage)	9	9	9	9	9	9	
<b>4ème code retour : contrôle "n° d'identification administratif de séjour"</b>							
Pas de problème à signaler	0	0	0	0	0	0	0
Non renseigné (longueur nulle/20 caractères espace)	1	1	1	1	1	1	
Traitement impossible (i.e. absence de génération d'une clé de chaînage)	9	9	9	9	9	9	
<b>5ème code retour : "fusion ANO-HOSP et HOSP-PMSI"</b>							
Pas de problème à signaler	0	0	0	0	0	0	0
N° Administratif dans fichier HOSP-PMSI mais pas dans le fichier de ANO-HOSP #	1	1	1	1	1	1	
Problème de calcul de clé d'intégrité sur ANO-HOSP	2	2	2	2	2	2	
Présence des 2 erreurs	3	3	3	3	3	3	
Traitement impossible	9	9	9	9	9	9	
<b>6ème code retour : "fusion ANO-PMSI et fichier PMSI"</b>							
Pas de problème à signaler	0	0	0	0	0	0	0
N° séquentiel PMSI dans fichier PMSI mais pas dans le fichier ANO-PMSI #	1	1	1	1	1	1	
Traitement impossible	9	9	9	9	9	9	
<b>7ème code retour : contrôle "date de référence"</b>							
Pas de problème à signaler		0	0	0	0	0	0
Erreur de format de la date de référence		1	1	1	1	1	
Date de référence fictive trop ancienne i.e. <1990		2	2				
Date de référence fictive trop ancienne i.e. <1940				2	2	2	
<b>8ème code retour : contrôle "Cohérence date naissance"</b>							
Pas de problème à signaler							0
Date de naissance du fichier de chaînage différente de celle du résumé PMSI							1
<b>9ème code retour : contrôle "Cohérence sexe"</b>							
Pas de problème à signaler							0
Code sexe du fichier de chaînage différent de celui du résumé PMSI							1

\* Pour SSR et RIM-P pas de changement par rapport à 2005

\* Optionnel pour SSR et RIM-P en 2009 (la distinction se fait en fonction du n° de format VID-HOSP indiqué dans le fichier de chaînage)

# Pour le RIM-P le code produit vaut dans ce cas 2 qui peut être distingué de "Problème de calcul de clé d'intégrité sur ANO-HOSP" pour lequel les autres codes retour ne valent pas 9

## **Travaux préparatoires : analyse de la cohérence des données PMSI**

Le mode de constitution de la base de données nationale des résumés PMSI chaînés, depuis la collecte d'information jusqu'à la consolidation, présente à priori les conditions pour être une source de description des parcours de soins hospitaliers : exhaustivité des établissements de santé et des séjours, uniformisation du traitement informatique de transmission des fichiers de résumés, en l'absence desquelles il conviendrait de s'attendre à des ruptures indues dans l'observation des parcours de soins et qu'il faudrait alors estimer. L'usage premier de la base de données nationale PMSI est centré sur le suivi de l'activité des établissements de santé et le pilotage de la valorisation financière de celle-ci. L'analyse des parcours de soins hospitaliers en est une utilisation potentielle et, pour la conforter, il a semblé souhaitable de conduire un ensemble d'analyses descriptives tendant à apprécier la cohérence effective de son contenu au regard de cette finalité.

### **Comparaison de résultats issus des données du PMSI à ceux d'autres sources**

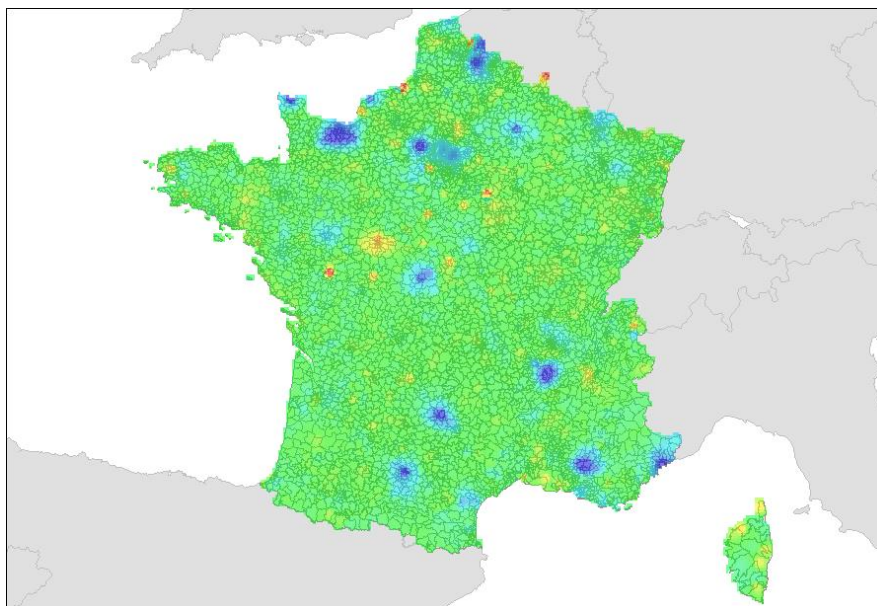
Une première approche est, pour des pathologies ou motifs d'hospitalisation pour lesquels un recueil systématique spécifique est réalisé, de comparer à cette référence, la description qu'il est possible d'obtenir à partir des résumés de la base de données nationale PMSI correspondant à cette pathologie ou motif d'hospitalisation ; eut égard à leur diversité il n'est pas envisagé de pouvoir couvrir l'ensemble du spectre des pathologies prise en charge en hospitalisation. Pour que cette comparaison soit aussi précise que possible, ont été considérées des pathologies avec prise en charge hospitalière exclusive ou presque, ayant une forte présomption de codage complet car impliquant directement la valorisation de l'activité de chaque établissement de santé, et pour lesquelles le même type d'agrégation des résultats que ceux disponibles pour la source alterne, pourra être obtenu dans le PMSI et ceci sans recours à des données de parcours hospitalier. Le résultat escompté de cette comparaison est d'apprécier en des termes généraux le biais éventuel dans les résultats issus des données du PMSI. Pour mémoire, la validation détaillée des données du PMSI nécessite une mise en œuvre spécifique avec un retour au dossier pour confronter le contenu des résumés PMSI aux documents du dossier médical hospitalier, ce qui n'est pas envisagé ici.

Quatre motifs d'hospitalisation ont été retenus : la naissance par comparaison avec l'état civil comme synthétisé par l'Institut national des statistiques et des études économiques ; les accouchements par comparaison avec l'enquête nationale périnatale ; les maladies à déclaration obligatoire par comparaison avec les données de veille sanitaire ; les prélèvements et transplantation d'organes par comparaison avec les données de coordination de l'Agence de biomédecine.

Pour les naissances, le critère appliqué aux données PMSI était volontairement très simple, un âge de 0 jour (i.e. la date d'entrée est celle de la naissance) et un mode d'entrée indiqué comme étant le domicile. Pour correspondre aux données issues de l'état civil, les résumés PMSI avec un code géographique de résidence de territoires d'outre-mer ou de pays étranger ou encore avec un code diagnostic principal de mort-né, ont été écartés. En 2008, 826772 naissances vivantes ont été recensées en France; dans le PMSI les critères appliqués donnent un nombre de 815994 soit 98,7% des données d'état-civil ; en 1999 le ratio était de 96,1%. Un décompte par code géographique a été réalisé et comparé au décompte des naissances par communes regroupées dans la mesure du possible selon le même niveau de découpage territorial. Pour les quelques communes correspondant à plus d'un code géographique, en l'absence de règle de répartition, les résumés ont été attribués au code géographique desservant le plus grand nombre d'habitants. L'écart médian, par code géographique, était d'une naissance en moins dans le PMSI avec un interquartile compris entre -3 et 1. La figure 1 présente cette comparaison territoriale avec un lissage spatial : en vert les zones avec des valeurs identiques, en bleu les zones avec un défaut des données PMSI, qui correspond à des établissements ne documentant pas l'information de résidence des nouveau-nés, en jaune et rouge un excès des données PMSI certains

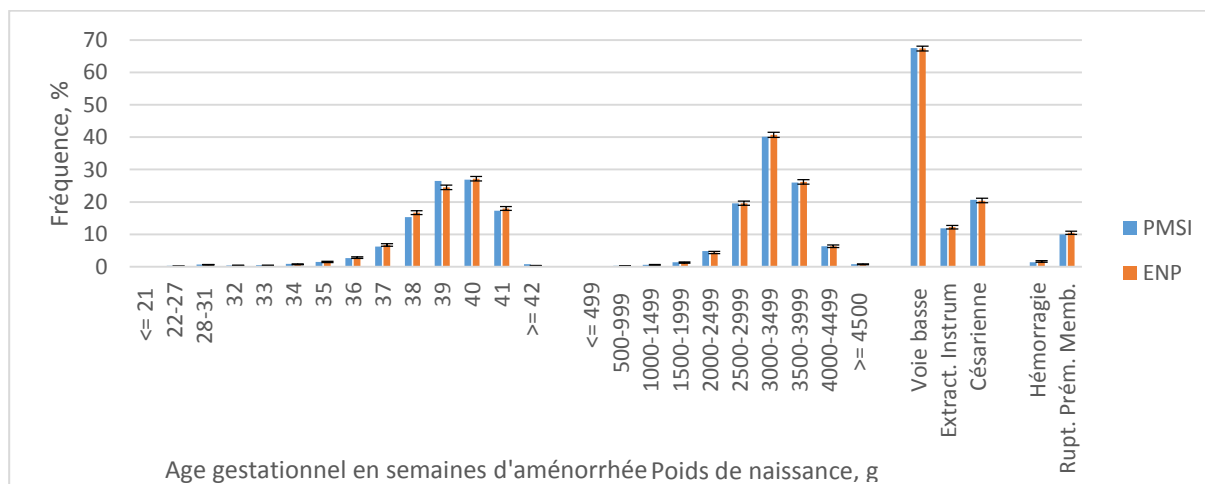
établissements semblent enregistrer par défaut dans le PMSI une date de naissance égale à la date d'admission, ce qui sera objectivé ultérieurement via le chaînage des résumés.

Figure 1 : Différence du nombre de naissance estimé avec le PMSI et celui donné par l'état-civil, France métropolitaine, 2008



Les caractéristiques disponibles dans les résumés PMSI correspondant à des accouchements en 2010 ont été comparées avec les résultats de l'enquête nationale périnatale (Quantin et al, 2014b). Il s'agit d'une enquête transversale auprès de tous les enfants nés en France une semaine donnée à intervalle régulier. En 2010, 823360 résumés d'accouchements et 829351 de naissances d'enfants vivants ont été sélectionnés dans la base de données du PMSI ; l'enquête nationale périnatale de 2010 comporte 14761 questionnaires relatifs à des naissances vivantes. Les distributions de l'âge gestationnel, des poids de naissance, des modes d'accouchement, ainsi que la fréquence des hémorragies de la grossesse comme de rupture prématurée des membranes apparaissent identiques, l'intervalle de confiance des estimations issues de l'échantillon de l'enquête comprenant ce que donne le PMSI excepté deux valeurs d'âge gestationnel (39 et 41 semaines) (figure 2, pour les données de France métropolitaine).

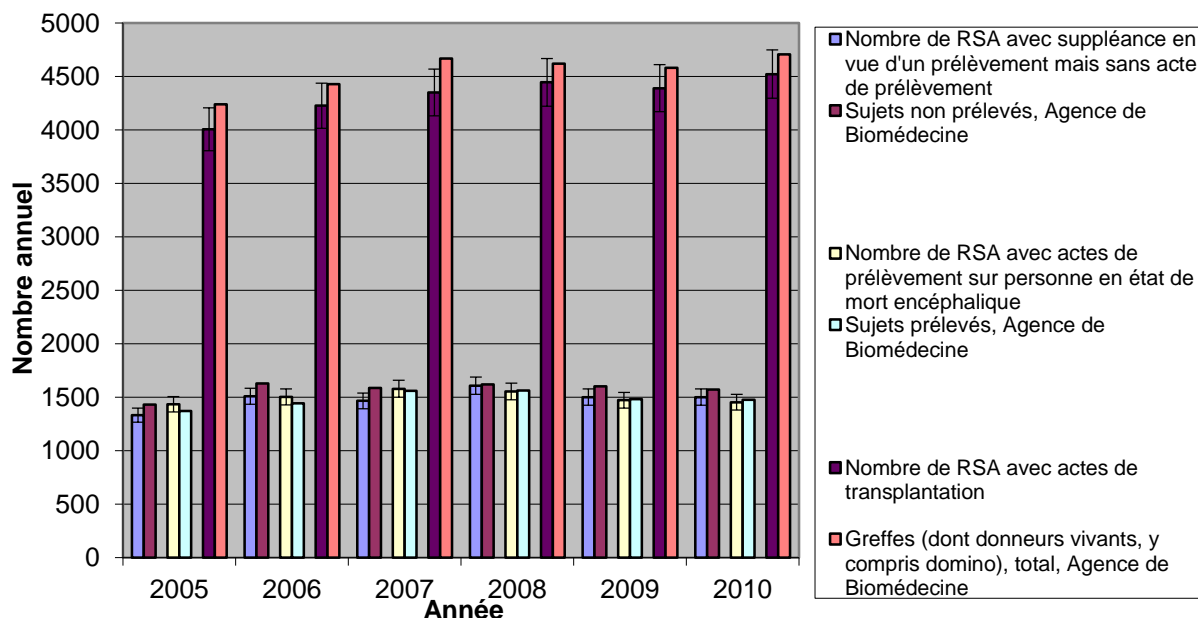
Figure 2 : Comparaison des caractéristiques des accouchements issues du PMSI avec celles issues de l'enquête nationale périnatale en 2010.



Un dénombrement des codes de diagnostic de maladies à déclaration obligatoire a été réalisé parmi l'ensemble des codes de diagnostic présents dans les résumés PMSI et le nombre de séjours correspondant à chacune de ces maladies comparé aux données de veille sanitaire pour chacune des années de la période 2000-2010. Les dénombrements issus du PMSI sont constamment supérieurs aux données de déclaration consultables sur le site internet de l'Institut de veille sanitaire, à l'exception de la rougeole et des toxi-infections alimentaires. L'utilisation de l'information de chaînage des résumés PMSI permet de dénombrer des nouveaux patients, ce qui fait alors coïncider les dénombrements entre les deux sources pour les infections invasives à méningocoque, les légionelloses, listérioses, tularémies. Bien qu'il n'y ait pas de contrôle spécifique de ces codes lors du traitement en routine du PMSI et donc un risque de codage par erreur, les données PMSI pourraient servir de point de départ pour interroger les établissements de santé, notamment pour des pathologies rares et sévères comme le tétanos, le botulisme, la brucellose. Les données du PMSI semblent compléter utilement les informations disponibles de veille sanitaire pour des infections plus fréquentes : tuberculose, hépatites A ou B aiguë, VIH, paludisme importé.

Enfin, l'activité de prélèvement et transplantation d'organes recueillie dans les résumés PMSI a été comparée à la synthèse des données de coordination issue du système d'information opérationnel dédié de l'Agence de biomédecine. Pour les années 2005 à 2010, le nombre de personnes prélevées, le nombre de personnes non prélevées, le nombre de transplantations ainsi que la nature de l'organe transplanté apparaissent identiques entre les deux sources à 5% près (figure 3). Ceci semble résulter du fait qu'il s'agit d'une activité hospitalière exclusive, avec une valorisation spécifique et devant être codée comme telle dans le PMSI. Contact a été pris avec l'Agence de biomédecine pour confirmer que les données de synthèse étaient bien issues du système de gestion en temps réel de cette agence.

Figure 3 : Comparaison entre PMSI et les données de l'Agence de biomédecine, du nombre de prélèvements et de celui des transplantations d'organes en France pour la période 2005 – 2010



Les résultats de cette première approche montrent la bonne exhaustivité de la base de données nationale des résumés PMSI, et ce avec un recul déjà important, ces deux caractéristiques constituant l'un des prérequis pour description des parcours de soins hospitaliers sans biais. Un autre prérequis

pour une utilisation de la base de données à cette fin, est la valeur opérationnelle des informations de chaînage.

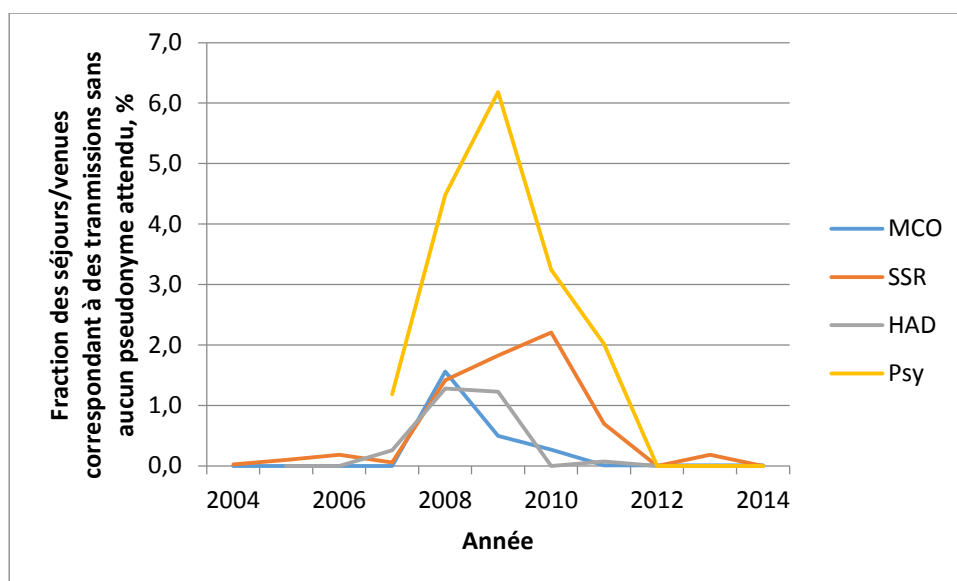
### **Description des pseudonymes dits « clés de chaînage » PMSI**

En préalable à l'évaluation de la valeur opérationnelle des informations de chaînage, la conformité de la génération informatique de celles-ci par les établissements de santé a été appréciée avec une série de vérifications indirectes reposant sur leur analyse descriptive détaillée dont les résultats sont mis en regard avec les spécifications techniques des méthodes retenues pour leur génération.

Les pseudonymes dits « clés de chaînage » dans la documentation afférente au PMSI, sont des chaînes de 16 caractères alphanumérique non signifiantes. Dans la base de données nationale, seul le résultat des tests sur les informations sources des pseudonymes, les qualifie d'un point de vue qualitatif ; le résultat de ces tests figurent dans une série d'indicateurs dites « codes retours ». La constance de ces indicateurs pour chaque pseudonyme a été mesurée pour ceux qui sont présents plus d'une fois (i.e. patients avec plus d'une hospitalisation) et est effectivement confirmée à des exceptions très ponctuelles près, qui sont la conséquence de l'évolution des tests réalisés par le logiciel incluant le module de génération des pseudonymes. Ainsi, des pseudonymes initialement qualifiés sans erreur dans les informations sources peuvent dans une année ultérieure être qualifiés autrement, les vérifications de format ayant été étendues. Ceci ne survient pas au cours d'une même année suite à la synchronisation de la mise à jour du logiciel dans les établissements de santé, tenus de télécharger la version de l'année depuis le site internet de l'Agence technique de l'information sur l'hospitalisation. Des combinaisons remarquables de codes retours existent : elles signifient alors que le triplet d'informations en entrée fait usage de la valeur conventionnelle pour désigner les informations manquantes (des « X », cf tableau 1). Pour ces combinaisons de codes retours, la constance du pseudonyme a été mesurée et effectivement constatée, à l'exception toutefois de la transmission de quelques établissements certaines années, transmissions dont il s'avèrera qu'elles comportent exclusivement des pseudonymes figurant dans aucune autre transmission de résumés PMSI.

Le constat précédent a conduit à développer une procédure d'analyse itérative. L'alimentation de la base de données nationale correspond in fine à une transmission par année et secteur d'activité de chaque établissement de santé, et à chacune de ces transmissions correspond une session particulière de génération de pseudonymes par mise en œuvre du logiciel installé dans les établissements. Afin d'apprécier si les pseudonymes figurant dans la base de données nationale PMSI avaient de fortes chances de d'avoir été générés selon la procédure attendue, tous les pseudonymes figurant dans une transmission comportant des pseudonymes identifiés par les codes retour remarquables, ont été considérés comme tels. Ce nouvel ensemble a servi ensuite de référentiel et appliqué à toutes les autres transmissions, lesquelles, dès qu'un de ces pseudonymes y figurait, impliquait que tous les autres avaient été générés de la façon attendue, ce qui étend alors d'autant le référentiel de pseudonymes. Le processus est poursuivi jusqu'à ce qu'aucune nouvelle transmission ne soit intégrée, soit après cinq étapes. Au décours des itérations, il est constaté que certaines transmissions n'ont pas été intégrées, donc apparemment avec exclusivement des pseudonymes générés de façon non attendue. Ces transmissions se révéleront en effet non intégrées dans le SNIIRAM car non connectable (communication des services de l'Assurance maladie). La figure 4 donne la proportion des séjours affectés par un défaut manifeste de génération, en fonction du secteur d'activité et l'année. Le corollaire est la démonstration que chaque établissement de santé a des patients en commun avec d'autres établissements de santé, aucune ne fonctionne de façon strictement isolée.

Figure 4 : Fréquence relative des séjours associés à des pseudonymes ne figurant dans aucune autre transmission, en fonction de l'année et du secteur d'activité

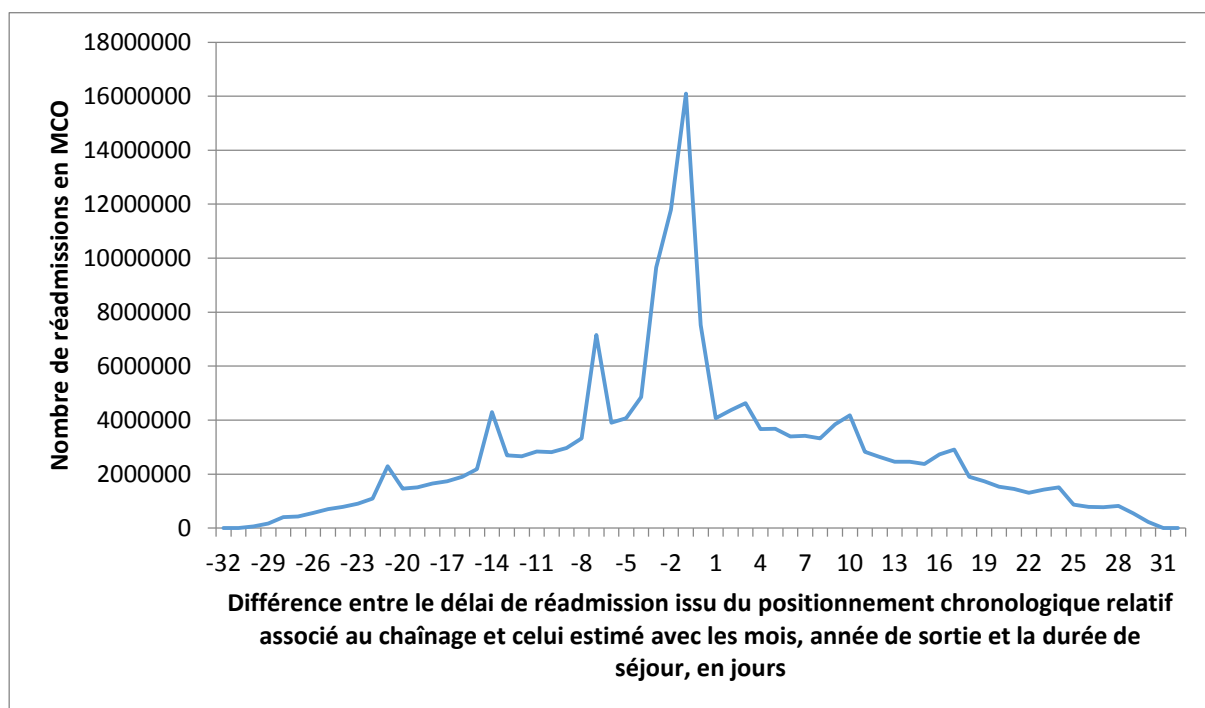


Le contenu des pseudonymes a ensuite été analysé afin d'apprécier la comptabilité avec les propriétés attendues des fonctions de hachage, SHA-1 en particulier : uniformité de distribution des caractères et absence de collision. Pour chacun des 16 caractères des pseudonymes, la fréquence d'apparition des chiffres 0 à 9 et lettres majuscules A à Z, non compris I, L, O, V (pour un alphabet de 32 caractères), sont observées avec des variations extrêmes inférieures à 0,4%. L'application d'une procédure itérative de troncature caractère par caractère des pseudonymes, pour faire apparaître des segments identiques, n'a d'effet que lorsque la longueur des pseudonymes réduite au moins de moitié, que la troncature soit appliquée à droite, à gauche ou alternativement. Pour ce qui est de l'absence de collision, aucun des 22 millions de pseudonymes distincts et sans anomalie signalée, figurant dans la base nationale PMSI de la période 2001-2003, n'est retrouvé ultérieurement. En effet, en 2004 le logiciel intégrant le module de génération des pseudonymes a été modifié afin que l'ordre des éléments de la date de naissance (jour, mois, année) dans le triplet d'information en entrée corresponde à ce qui était mis en œuvre par l'Assurance maladie pour constituer le SNiiRAM et permettre la connexion des résumés PMSI, Cette simple modification a pour effet qu'aucun pseudonymes d'une période n'est retrouvée dans l'autre alors qu'un certain nombre de patients ont été hospitalisés au cours des deux. Ce traitement a été effectué à l'Agence technique de l'information sur l'hospitalisation. Enfin, des séries de plusieurs millions de triplets d'informations ont été générés avec un format compatible et des pseudonymes générés avec le logiciel mis à disposition des établissements de santé, permettant de vérifier qu'à une information distincte en entrée correspond bien une unique en sortie. Pour mémoire, lors de simulation de grande ampleur, il convient de veiller à ne pas dépasser le nombre de valeurs distinctes que le générateur de nombres aléatoires est en mesure de produire.

La cohérence de l'information de positionnement chronologique relatif associée au pseudonyme a été testée vis-à-vis des mois et année de sortie figurant dans les résumés PMSI : les délais de réadmission i.e. hospitalisations successives d'un même pseudonyme, obtenus par chaque méthode ont été comparés ; pour rapprocher mois et année de sortie du positionnement chronologique relatif prenant appui sur la date d'entrée, la durée de séjour a été intégrée dans le calcul du délai de réadmission. Cette comparaison montre que la différence entre les deux méthodes est comprise dans la durée d'un mois, niveau le plus précis de positionnement chronologique des résumés PMSI sans les informations de chaînage mais obtenues par simple troncature de la date de sortie (i.e. sans traitement complexe). Cette constatation validation la cohérence des informations de positionnement chronologique, l'une des

composantes de la description des parcours de soins hospitaliers. La figure 5 donne la distribution observée pour les séjours e séances hospitaliers de court séjours de la période 2004-2014.

Figure 5 : Différence entre le délai de réadmission issu du positionnement chronologique relatif associé au chaînage et celui estimé à partir des informations des résumés PMSI de court séjour 2004-2014



En synthèse, les informations de chaînage à l'issue de ces vérifications de cohérence interne apparaissent générées conformément à ce qui est attendu, dans tous les établissements de santé et sur les années couvertes par le recueil PMSI, à de ponctuelles exceptions près. Dès lors, il est possible de prendre appui sur ces informations pour la description des parcours de soins. Deux sources d'information ont été utilisées pour évaluer la valeur opérationnelle des informations de chaînage des résumés PMSI pour cette finalité, le recueil expérimental du PMSI en psychiatrie, la base de données expérimentale Amphi.

### **Valeur opérationnelle des clés de chaînage PMSI comparée à celle de clés obtenues avec d'autres informations d'identification**

Un recueil expérimental de résumés PMSI relatif à l'activité de psychiatrie a été organisé par le ministère de la santé (Ministère de l'emploi et de la solidarité, 2002a) pour les années 2002-2005 auprès d'établissements volontaires. Dans ce recueil, l'activité de consultation réalisée dans les centres médico-psychologiques devait aussi être décrite et connectée le cas échéant à l'activité d'hospitalisation, à l'instar du système dit de la « fiche par patient » institué en 1988. Pour réaliser cette connexion, alors que les centres médico-psychologiques ne facturaient pas leur activité et ne disposaient donc pas du numéro d'assuré social des patients, il a été proposé de réaliser le chaînage des prises en charge avec un pseudonyme généré à partir d'un quadruplet d'informations en entrée : nom, prénom, date de naissance, code sexe. Pour les hospitalisations, le numéro d'assuré social étant disponible, le pseudonyme de chaînage PMSI était généré en parallèle. Il est dès lors possible d'analyser en situation réelle le devenir au fil des réadmissions de chaque couple de pseudonymes, ce qui est d'autant plus informatif que dans ce secteur d'activité les réadmissions peuvent être fréquentes.

Le logiciel générant le couple de pseudonymes, mis à disposition aux établissements qui en faisaient la demande par l'Agence technique de l'information sur l'hospitalisation, intègre le module FOIN de



l'Assurance maladie pour le pseudonyme PMSI et génère le second pseudonyme avec une méthode analogue : la fonction SHA-1 est appliquée au quadruplet d'information en entrée augmenté d'une information inconnue des utilisateurs ; le résultat est tronqué de moitié et transposé dans un alphabet de 32 caractères donnant un pseudonyme d'une longueur de 16 caractères. Des éléments de normalisation des informations nom et prénom ont été mis en place : longueurs respectives en pouvant excéder 50 et 30 caractères respectivement ; suivant la recommandation faite dans la documentation du logiciel, le nom doit être celui de naissance et le prénom le premier ; un traitement orthographique est appliqué au moment de la génération du pseudonyme : mise en majuscule, remplacement des caractères diacritiques par leur contrepartie de l'alphabet romain, suppression des caractères autres que les 26 lettres de l'alphabet et les 10 chiffres, en particulier les tirets, apostrophes, espaces, conversion le cas échéant vers l'encodage ANSI et transposition vers « l'endianess » attendue. Les données sont transmises semestriellement à l'Agence technique de l'information sur l'hospitalisation par les établissements qui participent à l'expérimentation. Des premiers résultats ont été présentés au congrès de l'Adelf. Une convention a été conclue entre le CHU de Dijon et l'Agence technique de l'information sur l'hospitalisation pour la mise à disposition de ces données aux fins de l'analyse des couples de pseudonymes de chaînage des résumés PMSI.

La base de données ainsi constituée comporte 0,5 million de séjours réalisés dans 122 établissements de santé, ce qui correspond à 67% de leur activité par comparaison avec la statistique d'activité des établissements et à 20% de l'activité d'hospitalisation de ce secteur pour la période considérée. Pour 315000 hospitalisations (63%), l'un des deux pseudonymes figure au moins deux fois, et peuvent alors être considérées comme des parcours avec réhospitalisation. Pour 11650 (3,7%) le pseudonyme PMSI n'est pas constant par rapport au pseudonyme issu du quadruplet comprenant le nom et prénom ; pour 11400 (3,6%) la situation est dans le sens inverse. Le pseudonyme PMSI apparait plus stable en cas de changement d'établissement et de façon générale pour les femmes. Le pseudonyme issu du quadruplet comprenant le nom et prénom apparait plus stable en cas de réadmission dans le même établissement et de façon générale pour les patients jeunes, âgés de moins de 21 ans. Si l'on considère chacune des 540 transmissions semestrielles, 13 d'entre elles comportent exclusivement des pseudonymes PMSI non retrouvées dans les autres, signant un défaut de génération des pseudonymes tant la situation est invraisemblable ; pour les pseudonymes issus du quadruplet comprenant le nom et prénom, ceci concerne 10 transmissions.

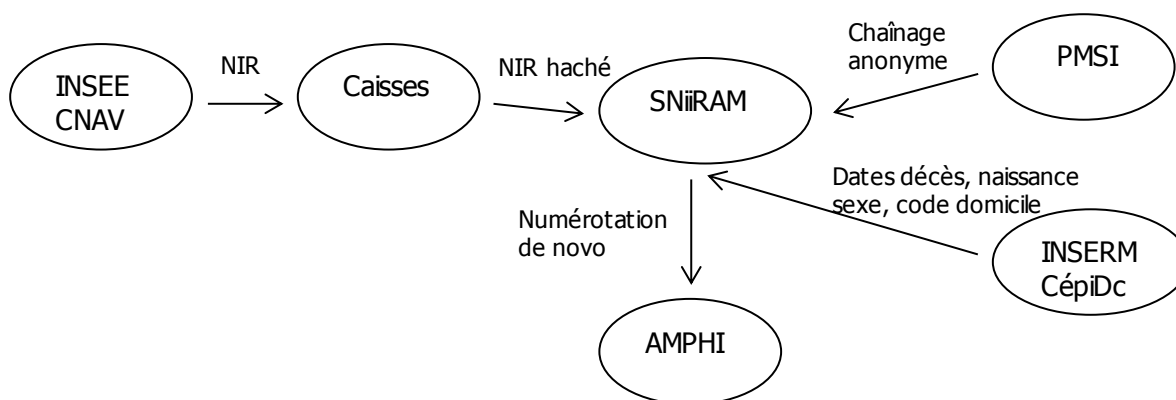
Dans cette situation de collecte rénovée de la description de l'activité d'hospitalisation en psychiatrie, bien que non encadrée par la valorisation financière, le pseudonyme PMSI apparait robuste, en particulier pour des parcours inter-hospitaliers. Par contre, et cela était attendu du fait de l'utilisation du numéro d'assuré social dans les informations source, les parcours de soins hospitaliers des enfants sont plus fréquemment affectés par des ruptures indues de chaînage. Malgré une mise en œuvre déterministe au moyen d'un pseudonyme global, un pseudonyme généré à partir du quadruplet nom, prénom, date de naissance, code sexe du patient peut constituer une alternative opérationnelle.

### **Apport de la base Amphi, précurseur du Système national des données de santé**

Dans la suite des travaux exploratoires initiés en 2004 sur la possibilité de mesurer et décrire les circonstances de la mortalité post-hospitalière en prenant appui sur le vecteur de connexion entre le PMSI et le SNNIRAM que constituent les pseudonymes du chaînage PMSI, la Direction de la recherche, des études, de l'évaluation et des statistiques, service statistiques du ministère de la Santé, a confié le soin à l'Institut national de la santé et de la recherche médicale et à la Caisse nationale de l'Assurance maladie de constituer pour son compte une base de données qui comprendrait en outre les informations médicales issues des certificats de décès. Le projet est baptisé l'Analyse de la Mortalité Post-Hospitalisation en France en 2008-2010 - recherche d'Indicateurs par établissement (Amphi). La base de données constituée comporte 24 millions de séjours correspondant aux données PMSI de

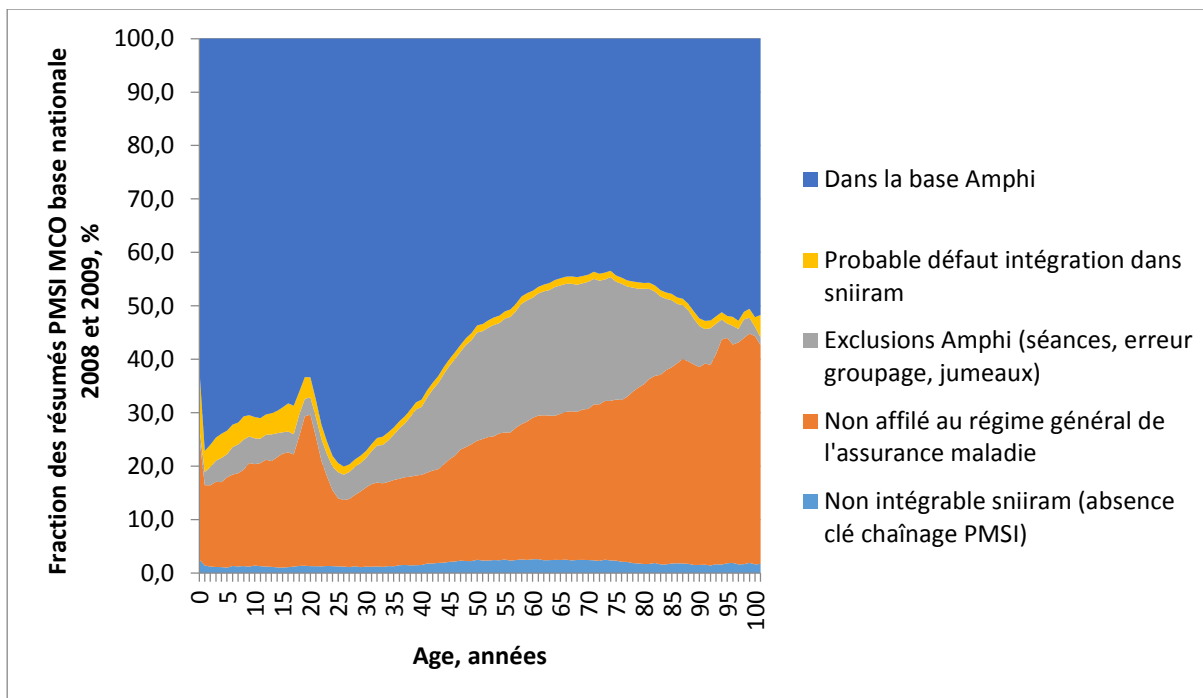
médecine, chirurgie, obstétrique des années 2008 et 2009 dont ont été écartés les résumés de séjour relatifs à des séances de traitement itératif, ou pour lesquels l’algorithme de groupage médico-économique a donné un résultat en erreur, ou encore pour lesquels un défaut dans les informations source du pseudonyme est mentionné. De plus, seuls les résumés PMSI des patients affiliés au régime général de l’Assurance maladie non compris les sections locales mutualistes (enseignants, ...) ont été pris en considération car l’information relative au statut vital n’était intégrée en routine dans le SNIIRAM que pour ceux-ci. Dans le SNIIRAM, par opposition au PMSI, les informations de datation figurent dans leur forme native à l’exception du jour de la date de naissance, ce qui a permis de réaliser l’appariement de la partie médicale des certificats de décès, cette section ne comportant pas d’information d’identité mais seulement des dates et lieux : dates de naissance et de décès, code postal du lieu de décès et du domicile, code sexe. Un appariement quasi exclusivement déterministe, i.e. exacte correspondance, sur ce profil d’informations a abouti pour 96% des décès survenant dans l’année qui suit la sortie d’hospitalisation alors même que le code de résidence n’est pas toujours univoque (postal ou Insee, patients avec plusieurs résidences). Cette méthode préfigure ce qui sera réalisé dans le cadre du système national des données de santé. La figure 6 présente l’articulation des différents appariements réalisés.

Figure 6 : appariements réalisés pour la constitution de la base de données Amphi.



La constitution de cette base de données a permis de mesurer, à travers les appariements réalisés, la qualité opérationnelle de deux informations du PMSI : le pseudonyme de chaînage des résumés et le mode de sortie « décès ». Par soustraction avec la base de données PMSI nationale de médecine, chirurgie, obstétrique pour les années 2008 et 2009, il a été possible d’estimer de façon globale à 2,1% la part des résumés PMSI, dont le pseudonyme ne trouve pas de contrepartie dans le SNIIRAM Cette estimation constitue une forme de validation externe de la valeur opérationnelle des pseudonymes générés dans les établissements de santé. Cette estimation ne comprend pas les séances de traitement itératif. La valeur opérationnelle des pseudonymes n’apparaît pas constante aux différents âges de la vie : pour les motifs mentionnés supra, le taux de défaut est plus élevé chez les moins de 20 ans (4%) et plus faible ensuite (1,4%) (figure 7, travail présenté au congrès de l’Adelf).

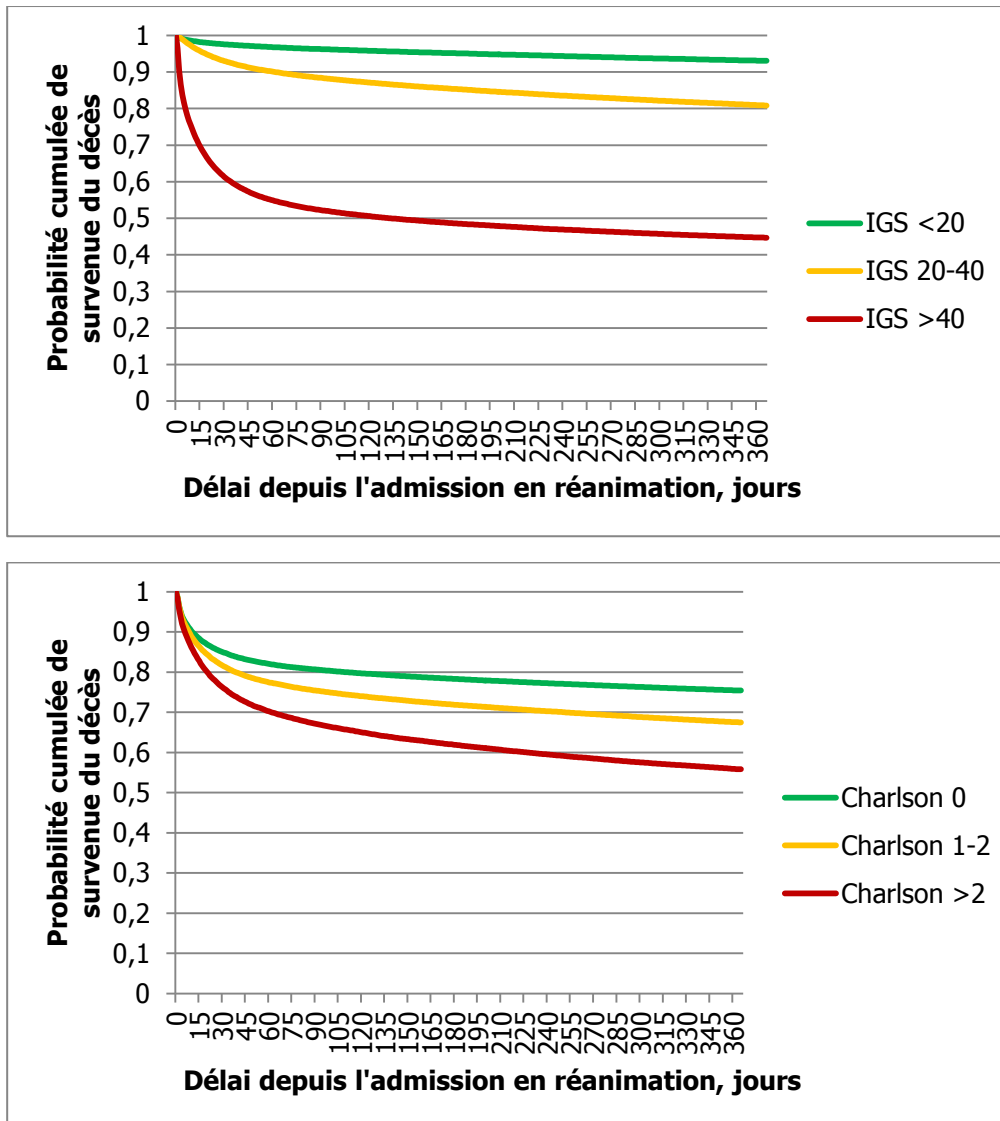
Figure 7 : répartition des résumés PMSI médecine, chirurgie, obstétrique des années 2008 et 2009 de la base de données nationale en fonction de leur appartenance à la base Amphi.



L'information « mode de sortie décès » des résumés PMSI a été comparée à l'information statut vital qui figure dans le SNIIRAM (enregistrement par les caisses primaires d'Assurance maladie lorsqu'elles sont directement informées, sinon transmission automatisée de cette information depuis la base de données nationale de l'état-civil). Cette mise en correspondance montre une concordance de 97%, dans 3% le décès est mentionné seulement sur l'une des deux sources. La discordance atteint 20% pour les enfants de moins de 1 an et diminue ensuite rapidement. L'information « mode de sortie décès » des résumés PMSI apparaît ainsi relativement fiable.

Afin d'apprécier la possibilité d'utiliser la base de données Amphi pour décrire le devenir des patients hospitalisés, il a été tenté de reproduire avec les données disponibles, les résultats d'une étude sur le devenir des patients ayant été admis en réanimation (Trouillet et al, 1996). La base de données Amphi comprend 290000 résumés de séjours avec admission en réanimation, ayant un score moyen de gravité à l'admission de 37 points et un score de Charlson non nul pour 38% le score étant recalculé à partir des codes diagnostics portés dans les résumés PMSI de ces séjours. Le score de Charlson identifie des pathologies chroniques ayant une incidence sur le pronostic vital dans un délai d'une année. Les courbes de survie selon ces deux facteurs sont présentées dans la figure 8, et montrent un résultat conforme à ce qui est attendu et publié : la sévérité à l'admission détermine la mortalité dans les premières semaines, à plus long terme la présence de pathologies chroniques est prépondérante.

Figure 8 : courbes de survie depuis l'admission en réanimation selon la sévérité à l'admission et la présence de pathologies chroniques – base Amphi



### Vérification formelle de la mise en œuvre de l'algorithme SHA-1

Les circonstances du développement et de mise en œuvre de la méthode de génération des clés de chaînage ont eu comme conséquence que seule une documentation technique extrêmement réduite a été diffusée. Il est seulement mentionné qu'une fonction de hachage est intégrée au dispositif, sans autre précision (Ministère de l'emploi et de la solidarité, 2001). Par ailleurs, la description initiale des tables constituant le SNIIR-AM, mentionne les fichiers PMSI comme un élément dans l'ensemble des fichiers de facturation (Ministère de l'emploi et de la solidarité, 2002b). Dans la base de données Amphi, il est ponctuellement constaté des clés de chaînage signalées comme ayant été générées à partir d'informations de format non conforme et pourtant issues du SNIIR-AM. Enfin, que la première étape de la fonction FOIN puisse être réversible était mentionné à différentes reprises (Fuchs, 2015, Zins, 2013).

Le centre de documentation de l'Assurance maladie, en réponse d'une demande d'informations techniques a communiqué un document qui indique que la fonction d'anonymisation est basée sur l'algorithme SHA (CESSI, 2002). Le centre d'étude des sécurités du système d'information de la caisse nationale de l'Assurance maladie, contacté ensuite, communique les informations complémentaires suivantes : l'algorithme SHA1 conforme à FIPS 180-1 a été appliqué dans la fonction FOIN ; une

troncature de la moitié de la signature produite par l'algorithme a été appliquée afin d'obtenir un résultat moins volumineux tout en restant suffisamment long pour éviter le risque de collisions indues ; la signature produite est transposée en un alphabet de 32 valeurs ; la fonction FOIN ajoute un caractère signant la version du secret ajouté aux données d'identification ; une moitié de ce secret est dans la fonction, l'autre dans un fichier extérieur, à produire au moment de l'utilisation ; l'implémentation est prévue robuste pour ne pas être affectée par le type de système informatique où la fonction sera mise en œuvre, little ou big endian, adressage 16, 32 ou 64 bits. Un test a été réalisé sur 30 millions de numéros de sécurité sociale distincts sans constater de collision, ceci au moment du développement de la fonction en 1996. Par contre, aucun document rendant compte de la mise en œuvre technique n'est communiqué et il ne semble pas qu'il y ait eu un audit externe, en particulier si les tests donnés dans FIPS 180-1 ont été reproduits.

Le fichier informatique de type exécutable (dynamic link library) comportant la fonction FOIN a été inspecté avec un logiciel de traitement de texte permettant un affichage du contenu sous une forme hexadécimale. Les cinq valeurs initiales définies pour l'algorithme SHA-1 dans la documentation FIPS 180-1 sont retrouvées. Il en est de même pour les quatre constantes, chacune apparaissant 20 fois, l'algorithme réalisant 80 tours de hachage. Ceci semble confirmer la présence de l'algorithme SHA-1 dans la fonction FOIN. Cependant, lors de cette inspection, de nombreuses autres fonctions apparaissent coexister. Une analyse statique du fichier informatique a été conduite afin de préciser le moment où l'algorithme SHA-1 est appliqué : description du contenu en prenant appui sur de la documentation sur le format des fichiers de type exécutables, application d'un logiciel transposant les codes binaires en programme assembleur afin d'en comparer le résultat à des exemples de programme écrit en assembleur. Ceci n'ayant pas abouti, une analyse dynamique est réalisée (suivi pas à pas dans une fenêtre du logiciel de transposition en assembleur) en donnant en entrée les informations d'identité codées comme manquantes (renseignées avec des « X », code hexadécimal 58) : le déroulé est conforme à la description FIPS 180-1, et le résultat est bien celui attendu pour les informations en entrée. Ceci valide de façon expérimentale la génération, dès la première étape, des clés de chaînage par hachage des informations d'identité selon l'algorithme SHA-1 et donc les propriétés qui lui sont associées, notamment l'irréversibilité des clés générées.

## **Description des parcours de soins hospitaliers avec le PMSI : matériel et méthodes**

La cohérence des informations de chaînage des résumés PMSI, pseudonyme, information de positionnement chronologique relatif, indicatrices de qualité des informations source du pseudonyme, ayant été confortées dans les étapes qui précèdent, cela ouvre la possibilité d'une utilisation de la base de données nationale du PMSI pour décrire les parcours de soins hospitaliers. Pour être pertinente en termes d'apport de connaissances à la pratique médicale ou sur le devenir des pathologies, le champ de la description envisagée doit être articulé avec un questionnement précis et en vérifiant que les informations de synthèse constituant les résumés PMSI seront à priori suffisantes pour apporter des éléments de réponse quantitatifs. Les caractéristiques du recueil PMSI, systématique (pas de réinterprétation à posteriori selon un canevas), prospectif (moins de manques), exhaustif (pas d'effet de sélection) au niveau national, permettent d'envisager des applications diverses, en terme de devenir et de caractéristiques associées en particulier. Un part de risque de travaux statistiques in fine non contributif est cependant à prendre en considération. Pour conduire de telles études, outre le canevas de travail, il convient de disposer d'un outil informatique adapté, d'une pratique confirmée de cet outil, d'une documentation technique détaillée des fichiers PMSI complétée le cas échéant d'informations permettant de restituer le lien avec l'activité clinique, d'une autorisation d'accès aux données. Ces points sont précisés dans ce qui suit.

Eu égard au nombre de résumés contenus dans la base de données nationale PMSI, plusieurs dizaines de millions, l'utilisation d'un système informatique paraît indispensable. Ce système doit pouvoir prendre en compte, sans saturation ni blocage, un tel nombre d'enregistrements informatiques. Le système doit pouvoir trier à façon les enregistrements, selon un nombre de critères dans l'idéal non limité. Dans le cas présent, la fonction de tri sert pour restituer l'ordre chronologique des résumés PMSI associés à un même pseudonyme, l'ordre des résumés étant par construction aléatoire dans la base de données nationale. Le recodage automatisé des informations doit pouvoir être fait selon différentes approches : application de conditions, mise en œuvre d'un référentiel externe, mise en relation entre résumés PMSI. Ces méthodes permettent de déduire de ce qui figure nativement dans la base de données PMSI, des informations nouvelles, utiles à l'analyse statistique, comme par exemple la durée entre deux séjours d'hospitalisation. De façon plus pragmatique, la possibilité de sélectionner un sous-ensemble de résumés, ou encore un sous ensemble d'informations, mais aussi de transposer des informations dans un format de taille minimale sans perte d'informativité, sont toutes des techniques qui permettent de limiter l'utilisation des ressources informatiques au strict nécessaire, et donc la durée des temps de calcul. Enfin, pour le calcul de statistiques descriptives comme pour l'application de modélisation statistique, le recours à des fonctions logicielles éprouvées évite un risque d'erreur inhérent à tout développement informatique.

Un tel système est disponible dans les sites où il a été accédé aux données PMSI pour réaliser les travaux dont les résultats sont présentés ici. L'intégralité des commandes informatiques employées pour mettre en forme les données et produire des descriptifs et modélisations statistiques, a été enregistrée au fur et à mesure dans un fichier de texte brut. Il n'a pas été fait usage d'interfaces graphiques de sélection et synthèse comme cela est parfois proposé. En effet, le fait de disposer de la séquence des commandes permet leur revue aux fins de vérification et, le cas échéant, d'apporter des corrections et modifications. De plus, certaines séquences peuvent être réutilisées dans des analyses de données ultérieures qui en sont ainsi facilitées car s'appuyant sur un canevas là aussi éprouvé. Ce formalisme permet aussi des échanges au sein des équipes de recherche. L'ajout de commentaires documentant chaque séquence de commandes facilite leur reprise, par exemple lors de questions transmises par les relecteurs des articles soumis en vue de leur publication. Toutefois, ceci nécessite de connaître, au moins dans ses grandes lignes, la syntaxe des commandes, cette syntaxe étant propre à chaque système informatique de traitement et d'analyse des données. S'il y a lieu d'approfondir ou d'étendre les méthodes employées,

de nombreuses ressources sont disponibles en ligne (documentation, exemples), identifiables par des moteurs de recherche sur internet.

En pratique, pour la restitution des parcours hospitaliers à partir des résumés PMSI (figure 9), après la jointure des fichiers comportant les informations de chaînage aux résumés correspondants (via champ et année d'activité, numéro Finess d'identification des établissements et numéro séquentiel PMSI), et l'ajout au positionnement chronologique du délai entre admission et début du résumé PMSI, un tri doit être réalisé sur les informations suivantes : pseudonyme, positionnement chronologique, durée couverte par le résumé PMSI, le mode d'entrée en ordre descendant, le mode de sortie ; les trois derniers critères sont destinés à la prise en compte de résumés de séjours d'une durée de 0 jours avec, en cas de réadmission ou transfert au cours de la même journée, un positionnement le plus plausible en termes de prise en charge, les critères de tri restituant une séquence dans l'ordre suivant : domicile (codé 8), puis transfert (codé 6 ou 7) enfin retour au domicile. Un fois l'ordre chronologique des résumés PMSI restitué par pseudonyme, les indices relatifs au parcours de soins hospitaliers peuvent être calculés : sont considérés comme transfert des résumés jointifs avec des numéros d'établissements successifs distincts et si le mode de sortie du premier résumé ou le mode d'entrée du second est codé comme tel, ce qui indique la poursuite de la prise en charge au cours d'un même épisode hospitalier ; les autres situations de résumés PMSI successifs pour un même pseudonyme sont considérées comme des réadmissions ; le délai de réadmission est calculé comme la différence entre la position chronologique relative du jour d'entrée du séjour suivant et la position chronologique relative du jour de sortie du séjour précédent ; si la réadmission a lieu le jour même le délai prend alors une valeur nulle ; des résumés de séjour peuvent parfois apparaître comme simultanés, couvrant en partie la même période ; ceci survient en cas de prestation inter-établissements, un séjour court inclus dans un séjour plus long, ou encore lors de traitements itératifs (chimiothérapie, radiothérapie, dialyse) poursuivi alors qu'il y a une hospitalisation dans un autre établissement ; très ponctuellement, des résumés apparaissent simultanés sur une longue période suite à une erreur dans l'enregistrement des dates de mouvement. Lors de la préparation du second travail publié présenté ici, la définition utilisée pour distinguer transfert de réadmission a été développée et appliquée pour distinguer poursuite de la prise en charge d'une tentative de suicide d'une réadmission pour une nouvelle tentative de suicide ; la clause « ou » est utilisée pour maximiser la prise en compte de la mention d'un transfert, sachant que la décision de transfert n'est pas toujours documentée comme telle dans les résumés PMSI, suite à une perte d'information au sein des établissements de santé entre équipes de soins et service en charge de la gestion administrative des patients. Lors de développement de méthodes de traitement automatisé de la base de données des résumés PMSI, chaque module de calcul d'indice a été vérifié au moyen de tests unitaires. De nombreux autres indices de parcours peuvent être utilisés comme par exemple, des éléments de parcours intra-hospitalier qui peut être une durée de séjour en soins critiques, un passage par l'unité d'hospitalisation d'un service des urgences, une durée de traitement (par exemple de ventilation mécanique) ; d'autres indices peuvent qualifier les réadmissions par exemple avec un diagnostic ou un acte particulier.

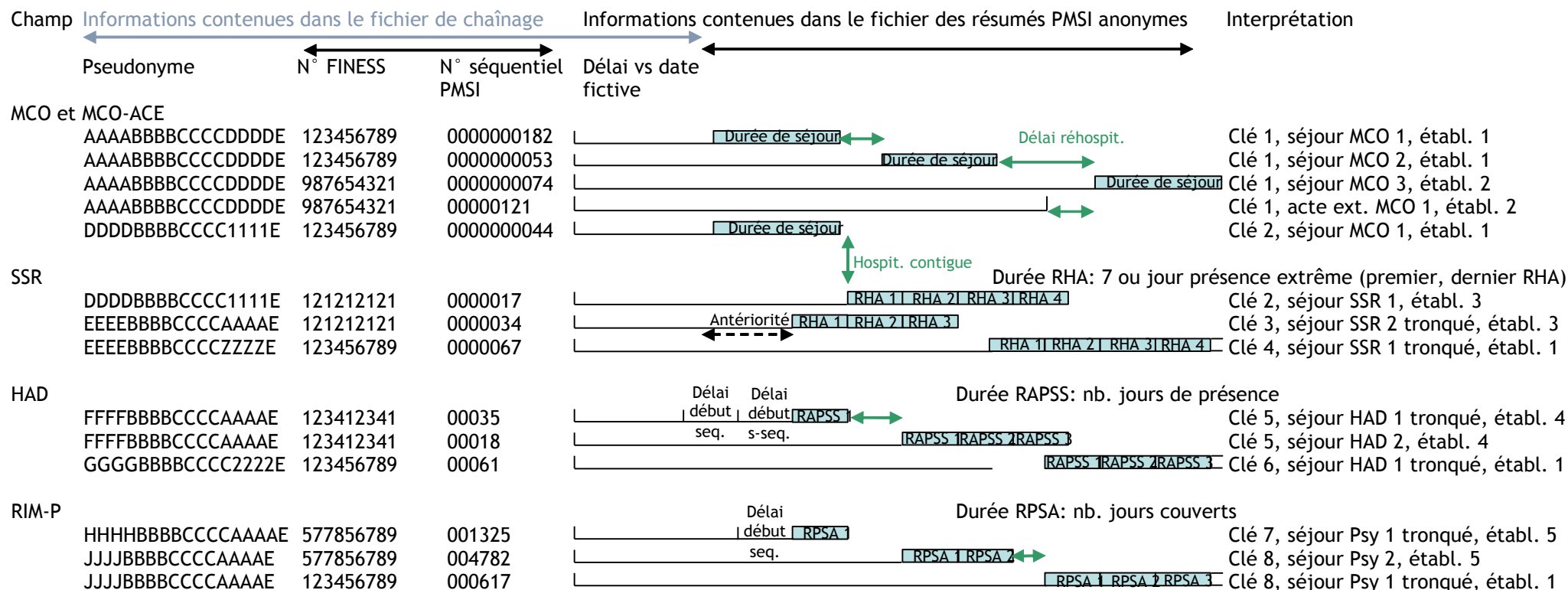
Pour le descriptif statistique, seuls les pseudonymes sans anomalie signalés dans les informations source (cf tableau 1, les codes retour doivent tous être de valeur nulle) et ce sur l'ensemble des séjours, sont pris en compte. La cohérence de l'âge, ou de l'année de naissance inférée, la constance du code sexe, sont vérifiés et le cas échéant les pseudonymes écartés (globalement <0,1%). Des informations issues de référentiels relatifs aux établissements de santé, par exemple le type d'établissement, la présence d'un service d'urgence, d'un bloc opératoire, le niveau d'une maternité, l'habilitation à faire des soins sans consentement, peuvent augmenter les possibilités de description et conforter l'exhaustivité et la cohérence des résultats obtenus. Ceci car l'on garde en mémoire que tout établissement peut avoir une organisation et une pratique propre pour ce qui concerne le recueil des informations des résumés PMSI. Il est donc vérifié que la prise en charge initiale des tentatives de suicide se fait bien dans des

établissements avec un service d'accueil des urgences ou de soins critiques et tous les établissements publics de santé ou d'intérêt collectif ayant de tels services prennent bien en charge des tentatives de suicide.

Les deux études présentées dans la suite du document ont été réalisées en suivant ce canevas méthodologique. L'accès aux données s'est fait dans le cadre réglementaire, autorisation permanente pour le service statistique du ministère de la Santé, autorisation spécifique délivrée par la Commission nationale de l'informatique et des libertés pour le service de biostatistique et d'informatique médicale du CHU de Dijon. Ces autorisations prévoient l'habilitation individuelle des personnes ayant accès, la mise en place de sécurités informatiques visant à se protéger des intrusions, et très récemment la journalisation des actions de chaque utilisateur.



Figure 9 : Diagramme illustrant le positionnement relatif des résumés PMSI anonymes, restitué à partir des informations de chaînage et de durées de séjour.



La jointure des informations du fichier de chaînage à celles du fichier PMSI, se fait sur le quadruplet champ, année, n° FINESS, n° séquentiel PMSI. Pour la clarté de la présentation, l'année n'est pas présentée ni les codes retours associés à chaque pseudonyme ; pour l'analyse des parcours, seuls les pseudonymes sans anomalie de génération sont pris en compte.

Avant de calculer des intervalles entre séjours d'un même patient, la base de données doit être triée le selon un ensemble d'information : pseudonyme, délai vis-à-vis de la date fictive propre à chaque pseudonyme, ce délai étant pour SSR, HAD, RIM-P augmenté du délai entre le résumé PMSI et le début du séjour, enfin, pour départager des mouvements ayant eu lieu la même journée, durée de résumé PMSI, mode d'entrée en ordre descendant, mode de sortie. Ce tri a pour effet d'ordonner chronologiquement les résumés PMSI comme présentée dans le diagramme. En vert sont indiqués les intervalles entre prise en charge et leur interprétation.

N°Finess : numéro d'identification des établissements de santé, information pivot pour la transmission des résumés PMSI ; n° séquentiel PMSI issu du réordonnement aléatoire ; MCO : médecine, chirurgie obstétrique (court séjour) ; MCO-ACE : actes et consultations externe des établissements de court séjour ; SSR : soins de suite et de réadaptation (moyen séjour) ; HAD : hospitalisation à domicile ; RIM-P : recueil d'information médicalisé pour la psychiatrie ; RHA : résumé hebdomadaire ; RAPSS : résumé par sous-séquence ; RPSA ; résumé de séquence

## Résultats : deux exemples d'application

Suivant cette méthodologie, deux études prenant appui sur le chaînage des résumés PMSI ont été conduites pour illustrer les possibilités de description et d'analyse des parcours de soins hospitaliers prenant appui sur la base de données nationale du programme de médicalisation du système d'information.

Une première étude s'attache au devenir dans les suites proches, le premier mois, des interventions chirurgicales. Le critère mesuré consiste en une information, la réadmission en court séjour. Le taux de réadmission dans les 30 jours est parfois considéré comme un indicateur reflétant la qualité des soins. L'analyse porte sur l'évolution de la fréquence de ces réadmissions sur une période de 11 années. À travers une série de descriptions et de modélisations prenant appui sur les informations disponibles dans les résumés PMSI, âge, sexe, groupe de pathologie, durée de séjour, type d'établissement de santé, il est tenté de discerner le facteur prépondérant dans l'évolution de ce taux, réforme du financement de l'activité de soins des établissements de santé ou modification des pratiques médicales.

La seconde étude saisit l'opportunité de données PMSI enrichies du statut vital dans l'année qui suit chaque hospitalisation, et en cas de décès de la cause médicale, informations obtenues par appariement aux données de l'Assurance maladie et de l'INSERM-CépiDc, pour décrire de façon quantifiée dans ses différents aspects le devenir des patients hospitalisés pour lésions auto-infligées non immédiatement fatales. Le mode de constitution de la base de données assurant la robustesse de l'information de statut vital dans l'année, ouvre la possibilité de quantifier des facteurs de risque de décès, et son étendue de quantifier des situations peu fréquentes.

## Évolution sur la période 2002-2012 du taux de réadmission après une hospitalisation pour chirurgie

Article publié dans British medical journal open, 2018

Résumé

Objectif

Une modification du mode de financement des établissements de santé peut induire une modification des pratiques de soins, afin de s'adapter à la nouvelle donne économique. La tarification à l'activité par forfaits peut constituer un incitatif à des prises en charge plus rapides pour permettre, à capacité égale, une plus grande activité. Cependant, ceci doit se faire sans dégradation de la qualité des soins. Un reflet indirect global de cette qualité peut être constitué par le taux de réadmission dans les 30 jours qui suivent un séjour pour chirurgie, ce type de soins hospitaliers étant leur ensemble plus normalisé que les séjours de médecine. En France, la tarification à l'activité des établissements de santé a été mise en place par palier entre 2005 et 2008, les établissements du secteur public ou privé à but non lucratif étant auparavant financés par un budget fixe et les établissements du secteur privé à but lucratif au prix de journée. Le lien entre cette réforme et le taux de réadmission dans les 30 jours suivant un séjour de chirurgie est étudié sur la période 2002 à 2012 avec les données de la base nationale PMSI.

Méthodes

Ont été sélectionnés dans la base de données nationale PMSI tous les séjours avec acte chirurgical dans la période 2002-2012 et ce pour les seuls établissements ayant une activité annuelle supérieure à 300 séjours de ce type. Ceci correspond, sur cette période de 11 années, à un ensemble de 51,6 millions de séjours effectués dans 740 établissements de santé en France. Pour chaque séjour, la survenue d'une

réhospitalisation en médecine, chirurgie, obstétrique dans les 30 jours suivant la sortie a été déterminée au moyen du chaînage des résumés. Les traitements itératifs, séances de chimiothérapie, radiothérapie, dialyse, n'ont pas été considérés comme des réadmissions. La fréquence de ces réadmissions a été modélisée avec une régression logistique et un modèle de séries temporelles avec seuils.

### Résultats

Entre 2002 et 2012, le taux de réadmission toutes causes confondues dans les 30 jours après séjour pour chirurgie augmente de 8,8% à 10,0% pour le secteur public ( $p < 0,001$ ) et de 5,9% à 8,6% pour le secteur privé ( $p < 0,001$ ). Sur cette période, l'augmentation du taux de réadmission dans les 30 jours après chirurgie est fonction du type de chirurgie réalisée. Une modélisation par des séries temporelles avec seuils montre une tendance linéaire générale prépondérante pour les deux secteurs. La mise en œuvre de la tarification à l'activité s'est accompagnée d'une modification significative de tendance pour le seul secteur privé ( $p < 0,001$ ).

### Conclusion

Sur la période à l'étude, il est observé une augmentation continue du taux de réadmission toutes causes dans les 30 jours suivants un séjour pour chirurgie, tant pour le secteur public que pour le secteur privé. Cette augmentation ne semble pas être associée à la mise en place de la tarification à l'activité mais semble refléter des modifications dans les pratiques devenant plus resserrées, avec une réduction progressive de l'étalement des prises en charge, pour la chirurgie de la cataracte et celle de l'implantation des chambres de perfusion notamment.

# BMJ Open Did case-based payment influence surgical readmission rates in France? A retrospective study

Albert Vuagnat,<sup>1,2</sup> Engin Yilmaz,<sup>2,3</sup> Adrien Roussot,<sup>1</sup> Victor Rodwin,<sup>4</sup> Maryse Gadreau,<sup>5</sup> Alain Bernard,<sup>6</sup> Catherine Creuzot-Garcher,<sup>7,8</sup> Catherine Quantin<sup>1,9,10</sup>

**To cite:** Vuagnat A, Yilmaz E, Roussot A, *et al.* Did case-based payment influence surgical readmission rates in France? A retrospective study. *BMJ Open* 2018;**8**:e018164. doi:10.1136/bmjopen-2017-018164

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-018164>).

Received 16 June 2017  
Revised 30 October 2017  
Accepted 19 December 2017

## ABSTRACT

**Objectives** To determine whether implementation of a case-based payment system changed all-cause readmission rates in the 30 days following discharge after surgery, we analysed all surgical procedures performed in all hospitals in France before (2002–2004), during (2005–2008) and after (2009–2012) its implementation.

**Setting** Our study is based on claims data for all surgical procedures performed in all acute care hospitals with >300 surgical admissions per year (740 hospitals) in France over 11 years (2002–2012; n=51.6 million admissions).

**Interventions** We analysed all-cause 30-day readmission rates after surgery using a logistic regression model and an interrupted time series analysis.

**Results** The overall 30-day all-cause readmission rate following discharge after surgery increased from 8.8% to 10.0% (P<0.001) for the public sector and from 5.9% to 8.6% (P<0.001) for the private sector. Interrupted time series models revealed a significant linear increase in readmission rates over the study period in all types of hospitals. However, the implementation of case-based payment was only associated with a significant increase in rehospitalisation rates for private hospitals (P<0.001).

**Conclusion** In France, the increase in the readmission rate appears to be relatively steady in both the private and public sector but appears not to have been affected by the introduction of a case-based payment system after accounting for changes in care practices in the public sector.

## INTRODUCTION

Financing hospitals is a challenge for any healthcare system. Many countries in the Organisation for Economic Cooperation and Development have chosen payment by diagnosis-related group (DRG). In 1983, the USA was the first country to introduce a case-based payment system according to DRGs of patients insured by Medicare.<sup>1–3</sup> Many countries around the world<sup>4</sup> chose to adopt this model as a tool to regulate hospital expenditure. The USA applied DRG-based reimbursement to one specific patient group, those 65 years and over (Medicare) and eventually for

## Strengths and limitations of this study

- To our knowledge, this is the first study to analyse 30-day all-cause readmission rates before, during and after the introduction of the case-based payment system in France.
- We linked individual patient data over 11 years for all surgical procedures performed in all acute care hospitals with >300 surgical admissions per year in France (n=51.6 million surgical admissions and 740 hospitals).
- We analysed rates of readmission for surgery with logistic regression models and with an interrupted time series analysis, in order to measure changes in readmission rates over time.
- One limitation of this study is that we considered all-cause readmissions as it is not possible to rule out planned readmissions in French claims data.

the poor (Medicaid). In other countries, only a part of hospital reimbursement is based on the DRG system, as in Portugal,<sup>5,6</sup> where this payment system concerns only certain care activities. To our knowledge, only France and Norway have implemented this case-based payment system to finance all hospital care activities since the early 2000s.<sup>7,8</sup>

Other countries, like Belgium, are considering the implementation of a similar case-based payment system,<sup>9</sup> but wonder whether it would induce certain unintended effects such as encouraging hospitals to increase their activity to improve their financial balance sheets. Moreover, whether or not there was improvement in quality of care<sup>10–12</sup> with regard to the decreased length of stays and in terms of mortality and readmission rates is a matter of debate.<sup>13–15</sup> Although hospital readmissions, when considered alone, can be used as an indirect marker of healthcare quality, their value in this setting is controversial.<sup>16,17</sup> In addition, there is some evidence that the implementation of a tariff system based on activity would lead to an increase in



For numbered affiliations see end of article.

**Correspondence to**  
Dr Catherine Quantin;  
[catherine.quantin@chu-dijon.fr](mailto:catherine.quantin@chu-dijon.fr)



rehospitalisation so as to maximise hospital revenues.<sup>18–21</sup> This effect was so feared in the USA and England that policy-makers imposed penalties for hospitals with abnormally high rehospitalisation rates.<sup>22–24</sup>

The medical information system in France has gathered exhaustive data on hospital activity since 1997, well before the implementation of case-based reimbursement in 2005. It is thus possible to obtain baseline rehospitalisation rates before the implementation of the case-based payment system. Since case-based payment was applied to all hospital activities, it is relatively easy to measure the evolution of readmissions after surgical procedures over the period of implementation.

The aim of this study is to determine whether implementation of case-based payment system was associated with a change in all-cause rehospitalisation rates in France. To do this, we compared rehospitalisations before the implementation of the case-based payment system (2002–2004), which was introduced stepwise in the middle of the study period (2005–2008) and after the implementation (2009–2012), after adjustment for the principal characteristics of patients. Previous studies conducted in France have not analysed the evolution of readmission rates over time<sup>25</sup> or only examined certain regions,<sup>13</sup> or were based only on specific diseases.<sup>26,27</sup> In this study, we include all surgical procedures and consider all readmissions, whatever the surgical subspecialty and cause of readmission.

## METHODS

### Source of data

The hospital discharge abstract database (Programme de Médicalisation des Systèmes d'Informations (PMSI)) contains individual, exhaustive and linkable but anonymous data on healthcare use for the whole French population and collects primary and associated diagnoses (secondary events and comorbidities) encoded using WHO International Statistical Classification of Diseases and Related Health Problems, 10th revision, and procedures performed during all hospital stays using the common classification system for medical procedures (Classification commune des actes médicaux). The very good quality of the French hospital database has previously been evaluated and has enabled us to carry out several epidemiological and health services research studies concerning hospitalised patients in France.<sup>25,26,28–31</sup> The study was approved by the National Committee for Data Protection.

### Population

This study was a retrospective multicentre study based on nationwide PMSI data.

We include all patients admitted to all acute care hospitals with surgical wards (740 hospitals including 295 public hospitals and 445 private hospitals) for surgical procedures (as defined by the French DRG classification) over 11 years (2002–2012). Hospitals with fewer than 300 surgical admissions per year were not included, because

many of them closed during the study period. We considered separately public and private hospitals, as hospital funding was completely different between these two types before the introduction of case-based payment in all hospitals. The 46 private not-for-profit hospitals were classified in the public sector, as their hospital funding was the same as for public hospitals.

### Main outcome measure: readmission within 30 days following discharge

For each selected surgery admission, the time from patient discharge to a new admission was calculated according to the linked information. Initial hospitalisations and stays ending in death or transfer, iterative treatments and neonatology were excluded. In 'iterative treatments', we considered 1 day admissions for treatments such as chemotherapy, radiation therapy and haemodialysis. All-cause readmission was defined as 'a new hospitalization within 30 days<sup>30</sup> following discharge after an admission for surgery, whatever the reason for this second admission' as done before,<sup>25,26</sup> that is, if a patient was readmitted for a reason other than the diagnosis for the first admission, it was still considered a readmission. The hospital where the readmission took place was also noted.

### Variables studied: characteristics related to readmission

The characteristics of the admissions were studied according to the variables available in the national medical-administrative database, namely year of hospitalisation, age, gender, mode of admission (from home, via an emergency service and transfer), the type of hospital, morbidity (Charlson score, Major Diagnostic Categories of French classification in DRGs that we called DRG groups) and length of stay.<sup>15</sup> We also added the urban/rural classification of patients' place of residence according to the French institute of statistics and censuses. We subdivided information regarding urban areas into three categories: city centres, suburbs of big cities and small towns.

### Statistical analysis

In the first analysis, we studied the influence of the variables defined above (all dichotomised) on readmission at 30 days with two logistic regression models. The probability of readmission was analysed separately for the two types of hospital sector (ie, public and private). The first model (M0) concerned all hospital admissions for surgery.

The second model (M1) excluded DRG groups with low volumes of activity (burns, infectious diseases, HIV diseases, multiple trauma, psychiatry in acute care and other types of care). They also excluded cases with major modifications in care practices during the period, either for changes in care management (eg, in ophthalmology) or therapeutic changes for the treatment of HIV. Regarding ophthalmological surgery, since cataract surgery is more and more frequently performed to one eye and rapidly after to the other (<1 month after), we had to take into account this change with time, which

results in an increase in readmission rates in ophthalmology substantially greater than in other specialties.

In the second analysis, an interrupted time series analysis was performed to measure changes in the readmission rate over time while taking into account the variables defined above. This model used monthly readmission rates over the study period and included a linear time trend. Three periods were considered: the pre-case-based payment system period (from 2002 to 2004), the implementation period (from 2005 to 2008) and the postimplementation period (from 2009 to 2012). In accordance with seasonal fluctuations, random error was modelled by an autoregressive model with a parameter at lag 12.

We thus quantified the impact of the implementation as changes in the level and slope compared with the preimplementation period. SAS V.9.4 was used for all of the analyses. The threshold of statistical significance was set at  $P < 0.05$ .

## RESULTS

### Descriptive study

The study sample contained almost 52 million admissions, accounting for all admissions with DRGs related to surgery in hospitals with  $>300$  admissions per year. Admissions with in-hospital deaths or without linkage information were excluded and represented  $<5\%$  of our admissions.

The number of admissions with surgery selected in the database increased from 4.1 million in 2002 to 5.3 million in 2012, for a total of 51.6 million admissions over the 11 years (table 1). Of the surgeries, 60% and 40% took place in profit-making private and in public or non-profit-making private hospitals, respectively. During the study period, there was a steady increase in the mean age of patients (from 48.6 to 51.3 years) and a decrease in the mean length of stay (from 4.3 to 3.0 days). The disease profile remained relatively stable, except for a slight increase in admissions in ophthalmology units.

Between 2002 and 2012, the readmission rate following admissions for surgery (figure 1) increased in both the public and private sectors: from 8.8% to 10.0% and 5.9% to 8.6%, respectively. Although the overall readmission rate was higher in public than in private hospitals ( $P < 0.001$ ), its increase appeared to be relatively steady in both sectors. However, this increase was significantly greater in the private than in the public sector ( $P < 0.001$ ).

The descriptive results underlined the disparity in readmission rates at 30 days between the different DRG groups over the study period (figure 2), in terms of both volume and evolution. In 2012, the readmission rate ranged from 2.7% for ear, nose and throat surgery to 26% for haematology and 27% for the surgical treatment of burns. Two types of surgery in particular showed a major change in the readmission rate: ophthalmology and HIV-related surgery. For ophthalmology, the readmission rate increased from 9.3% in 2002 to 16.5% in 2012 in the public sector and from 10.0% to 19.7% in the

private sector. For HIV-related surgery, the readmission rate in the public sector fell from 31.4% in 2002 to 25.4% in 2012, but peaked at 39.3% in 2006, with major variations from one year to another.

The profile for the evolution of readmission rates by type of surgery also differed according to the type of hospital and surgery (figure 2). For example, the increase in the readmission rate for ophthalmology was particularly pronounced in private hospitals, rising from 10.0% in 2002 to 19.7% in 2012. Concerning other types of surgery, the readmission rate for the public and private sectors remained quite stable.

### Multivariate models: study of factors associated with readmission

After adjustment for the DRG groups and morbidity, the probability of readmission at 30 days increased with time (table 2, model M0) in both the public and private sectors. We can see that the effect of the risk of readmission also increased with age and that this effect was greater in the private than in the public sector (eg, for patients aged 80 years and over,  $OR = 1.9$  in the public sector vs 5.3 in the private sector). Moreover, patients living in urban areas were slightly more at risk of readmission, with a more marked risk in small towns.

However, after excluding cases with major modifications in care practices during the period (such as ophthalmological surgery) or with low volumes of activity, the overall increase in the readmission rate found in model M0 was not retrieved for public hospitals and the readmission rate did not seem to increase with time after the implementation of the case-based payment (model M1).

### Interrupted time series model

The series exhibited significant linear trends over the period (see figure 3). Rehospitalisation rates increased by 0.0170 percentage points per month in public hospitals ( $P < 0.05$ ) and by 0.0224 percentage points per month in private hospitals ( $P < 0.001$ ). However, the implementation of case-based payment was associated with a significant increase in rehospitalisation rates for private hospitals ( $P < 0.001$ ).

## DISCUSSION

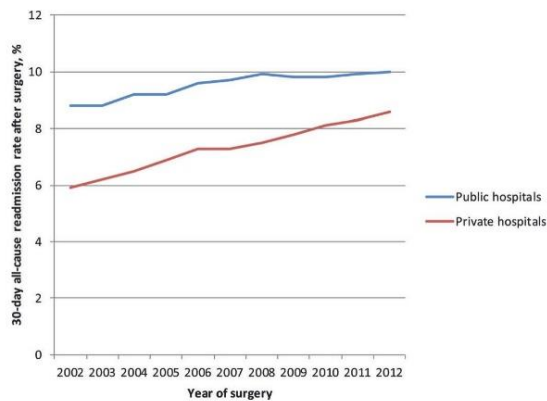
Our nationwide population-based analysis of 51.6 million hospital admissions for surgery over the 2002–2012 period found that the overall readmission rate within 30 days following discharge increased with time both in the public and private sectors, after adjustment for age, gender and comorbidities. The increase was greater in the private sector than in the public sector. However, after excluding cases with major modifications in care practices during the period, such as ophthalmological surgery, the overall increase in the readmission rate found in the previous regression logistic model was not retrieved and, for public hospitals, the readmission rate did not seem to have been influenced by the implementation of



**Table 1** Characteristics of patients and admissions, all surgical procedures (France, 2002–2012)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Age, mean (years)	48.6	49.2	49.5	49.8	49.7	49.8	50.3	50.6	51.0	51.0	51.3
Length of stay, mean (days)	4.3	4.2	4.0	3.8	3.7	3.6	3.5	3.4	3.3	3.1	3.0
Gender, male (%)	46.9	47.0	47.0	47.2	47.6	48.0	48.0	48.0	48.0	48.0	47.7
Type of hospital, admission (%)											
Private	60.2	61.3	60.1	60.5	59.9	59.1	58.2	58.4	58.2	58.2	58.6
Public	39.8	38.7	39.9	39.5	40.1	40.9	41.8	41.6	41.8	41.8	41.4
Admission through emergency department (%)	–	–	–	–	–	–	9.0	10.1	10.6	11.1	11.4
Groups of surgical diagnosis-related groups (%)											
Orthopaedics, rheumatology	26.4	26.2	26.5	27.0	27.3	27.4	27.6	27.8	27.7	27.6	27.4
Ophthalmology	11.4	12.2	12.5	12.7	12.9	13.0	13.2	13.7	14.0	14.2	14.5
ENT, stomatology	12.8	12.7	12.3	12.4	13.2	13.2	12.9	12.7	12.7	12.8	12.8
Abdominal	13.0	12.9	12.4	12.1	11.8	11.7	11.7	11.5	11.6	11.4	11.2
Gynaecology	9.3	9.3	8.7	8.4	8.5	8.3	8.2	8.1	7.8	7.8	8.0
Urology	6.8	6.8	6.8	6.9	7.1	7.3	7.3	7.4	7.6	7.7	7.7
Skin	5.9	5.8	6.6	6.8	6.1	5.9	5.9	6.1	6.2	6.4	6.4
Vascular peripheral disease	5.4	5.3	5.2	5.1	5.2	5.1	5.1	5.0	4.8	4.7	4.7
Nervous system	2.9	2.8	3.0	2.8	2.4	2.5	2.5	2.5	2.6	2.5	2.4
Cardiology	2.0	2.0	1.9	1.9	1.9	1.9	2.0	2.0	2.0	1.9	1.9
Endocrinology	1.2	1.1	1.2	1.1	1.1	1.1	1.1	1.1	1.0	1.0	1.0
Other	1.0	1.1	1.0	1.0	0.9	0.8	0.7	0.7	0.6	0.6	0.6
Pneumology	0.6	0.6	0.5	0.5	0.5	0.5	0.6	0.6	0.5	0.5	0.5
Haematology	0.7	0.7	0.7	0.7	0.6	0.6	0.6	0.6	0.5	0.5	0.5
Burns	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Severe trauma	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Infectious diseases (HIV excluded)	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.0	0.0	0.0	0.0
Psychiatry, suicide attempts	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Patients with HIV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Number of admissions	4058201	4143632	4322156	4529058	4639829	4722789	4806150	4921823	5017772	5186634	5270938

ENT, ear, nose and throat.



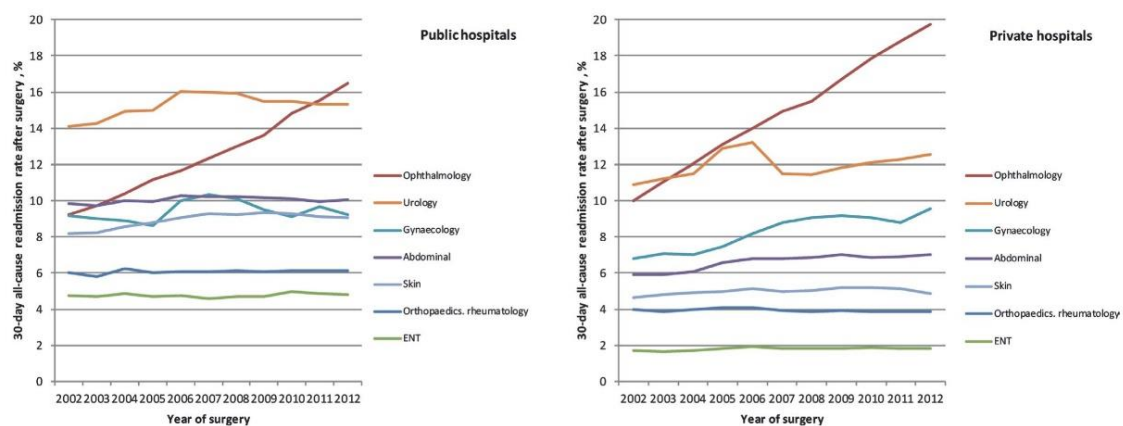
**Figure 1** Thirty-day all-cause readmission rates after surgery according to hospital sector, all surgical procedures (France, 2002–2012).

case-based payment. The interrupted time series analysis confirmed that the implementation of case-based payment was only associated with a significant increase in rehospitalisation rates for private hospitals. These results suggest that hospital reimbursement is not the only determinant of readmission.

These findings contradict the results of a retrospective observational study in the USA,<sup>32</sup> which found a decreased 30-day readmission rate following inpatient discharge for nine surgical specialties in the Veterans Health Administration (VHA) over a similar 10-year period (2001–2010). The fact that in France, no penalty is risked by hospitals in case of increased readmission rate may partially explain this difference. Moreover, our study included all types of surgery and specialties, including ophthalmology. We also considered all readmissions, whatever the sector, in contrast with the VHA study, in which patients having surgery at a VHA facility and then readmitted in the private sector could not be captured. In another study,

comparing patients insured by Medicare before and after the implementation of the case-based payment system,<sup>33</sup> the authors found that case-based payment was accompanied by a reduction in the length of stay. In parallel, the discharge mortality rate and the readmission rate did not increase. The same results were found by Kahn *et al*<sup>13</sup> with a 24% decrease in the length of stay and an unchanged readmission rate. Another early study on the effects of implementing Medicare in the USA reported stable in-hospital mortality rates and care quality.<sup>34</sup> At the same time, this stability of in-hospital mortality was put into perspective by Sager *et al*, who reported a significant rise in mortality at home and thus concluded that in-hospital deaths had been converted to at-home deaths in patients not covered by the new system.<sup>35</sup> In Europe, it is difficult to say whether mortality rates have been affected by implementation of the case-based payment system. Studies have nonetheless shown that these systems are often accompanied by shorter lengths of stay and an increase in the number of admissions and in hospital productivity.<sup>5,6,20,36</sup> Cutler hypothesised that payment linked to activity could have influenced the readmission rate, given that these rates increased in hospitals with deficits and thus under financial pressure.<sup>14</sup>

The evolution of readmission rates was slightly different in the public and private sectors. In France, the former generally manages the most complex cases of each disease, including emergency cases.<sup>37</sup> It is therefore not surprising to see a higher overall rate of readmissions in public than in private hospitals. However, comparison of the two sectors showed that the management of cataract surgery was reorganised faster in the private sector. The greater increase in readmissions in the private sector than in the public sector may be surprising, since the new pricing policy provided the least incentive for change in the private sector. The pricing policy before the case-based payment system already included payment according to activity in the private sector and readmissions were



**Figure 2** Thirty-day all-cause readmission rates after surgery according to the most frequent DRG groups, by hospital sector, all surgical procedures (France, 2002–2012). DRG, diagnosis-related group; ENT, ear, nose and throat.



**Table 2** Multiple logistic regression of 30-day all-cause readmission rates according to characteristics of patients and admissions, all surgical procedures (France, 2002–2012)

	Public hospitals (ORs)		Private hospitals (ORs)	
	Model 0	Model 1	Model 0	Model 1
<b>Year of surgery</b>				
2002	Ref.	Ref.	Ref.	Ref.
2003	0.994	0.984**	1.023**	1.000
2004	1.043**	1.039**	1.074**	1.036**
2005	1.033**	1.018**	1.152**	1.096**
2006	1.084**	1.070**	1.212**	1.147**
2007	1.093**	1.075**	1.223**	1.121**
2008	1.105**	1.077**	1.246**	1.127**
2009	1.091**	1.059**	1.305**	1.151**
2010	1.090**	1.040**	1.351**	1.155**
2011	1.103**	1.045**	1.395**	1.166**
2012	1.103**	1.033**	1.448**	1.191**
<b>Comorbidity</b>				
Charlson index (>0 vs =0)	1.943**	2.061**	1.529**	1.812**
<b>Admission</b>				
Home versus transfer from hospital	0.899**	0.850**	0.640**	0.613**
<b>Gender</b>				
Male versus female	1.096**	1.106**	1.024**	1.049**
<b>Age</b>				
<10 years	Ref.	Ref.	Ref.	Ref.
10–19 years	0.918*	1.010*	1.438**	1.404**
20–29 years	1.112**	1.274**	2.636**	2.592**
30–39 years	1.400**	1.624**	3.692**	3.650**
40–49 years	1.398**	1.599**	3.544**	3.401**
50–59 years	1.615**	1.850**	4.150**	3.869**
60–69 years	1.712**	1.962**	4.567**	4.142**
70–79 years	1.777**	2.009**	5.028**	4.577**
80 years and over	1.954**	2.263**	5.304**	5.433**
<b>Place of residence</b>				
City centre	1.004*	0.998	1.025**	1.032**
Suburbs	1.018**	1.008**	1.017**	1.019**
Small town	1.021**	1.011**	1.025**	1.002
<b>Fixed effects for each DRG group†</b>				
Interaction term: DRG group×year	No	No	No	No
Number of observations	20 893 246	18 036 369	30 459 905	24 736 141
<b>Concordance statistic</b>				
Concordant pairs (%)	66.7	66.2	71.4	69.9

\*\*P<0.01.

\*P<0.10.

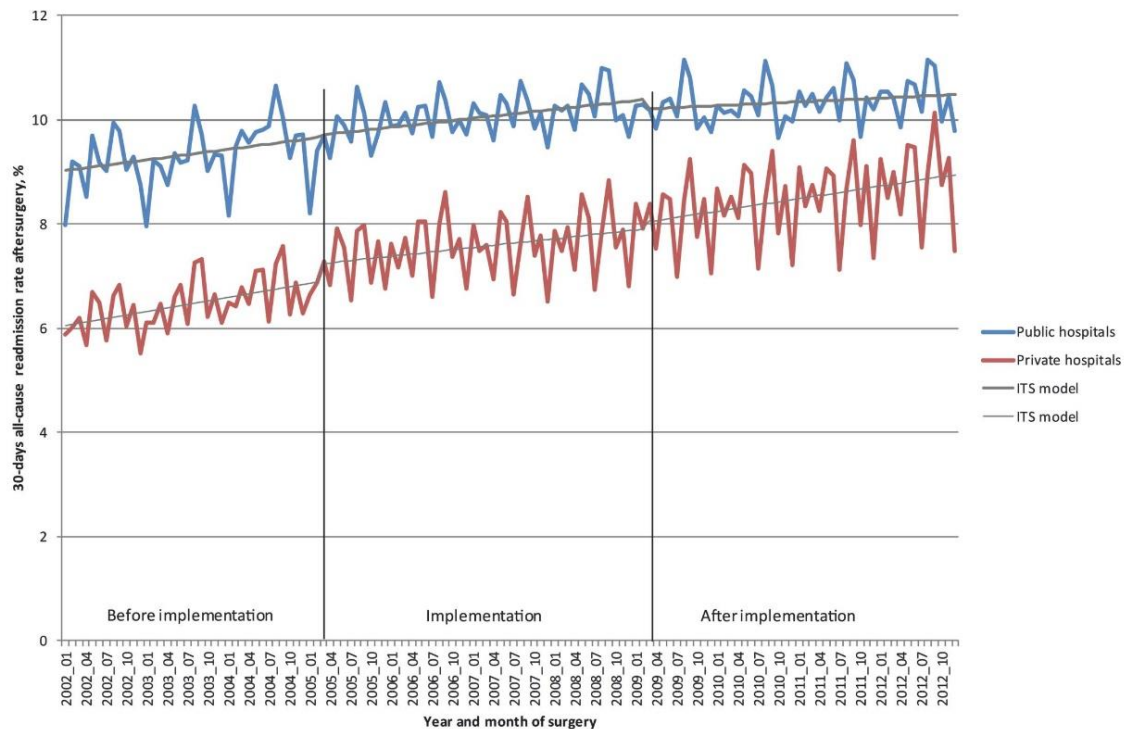
†French classification of DRGs.

DRG, diagnosis-related group.

already paid for before implementation of the case-based payment system.

As this rise in readmissions did not seem to be only related to the pricing reform, one might wonder whether

it was also related to changes in care practices. A more specific analysis of our results did not support this hypothesis. Two contrasting examples show the effect of changes in care practices on readmission rates. First consider the



**Figure 3** Global trends in 30-day all-cause readmission rates per month after surgery (France, 2002–2012): ITS analysis. ITS, interrupted time series.

case of cataract surgery, nearly 500 000 surgeries per year in France. These procedures have moved from inpatient to outpatient hospitalisation with prompt recovery leading to a shortened delay between surgeries for each eye. Consequently, their increased readmission rates only reflect this shortened delay between surgeries for each eye due to the improvement in practices and not a secondary deleterious influence of hospital funding. Second, in HIV-related surgery, we observed changes in the opposite direction, with a decrease in the readmission rate, which may only reflect the improved efficacy of antiretroviral treatments leading to fewer recurrent hospitalisations. These observations suggest that to interpret these results, all changes (population, clinical practices and payment incentives) need to be considered for each group of diseases.

At the international level, the financial impact of readmissions to hospitals has led to the implementation of different policies aiming to limit such admissions as much as possible. The impact of these measures has been investigated in American studies showing that the decrease in the number of readmissions in the population studied did not stem from the implementation of such policies, but rather from the long-standing adaptation of practices of healthcare staff, as shown in our study.<sup>38,39</sup> These results showed that an overall decrease in readmissions

at 30 days has to be considered over the long term rather than as a direct and immediate result of healthcare policy. A secondary effect such as a concomitant increase in outpatient consultations needs to be considered as well.<sup>40</sup> However, a recent study reported significant effects of such incentives, leading to decreases in readmission rates in small public-sector hospitals located in rural areas.<sup>39</sup> In our study, we considered the place of residence of patients and not the location of the hospital as in France most hospitals are located in urban areas. We only found a slight effect of the patients' place of residence on readmissions. We do not think that this result can be affected by the risk of ecological fallacy as we only included one aggregated variable in our logistic regression model.<sup>41</sup>

In the USA, some hospitals regularly publish their 30-day readmission rates with regard to cardiovascular or pulmonary diseases. However, a recent analysis of factors associated with readmission conducted in a cohort of patients insured by Medicare showed that not all hospitals were equally affected by readmissions.<sup>42</sup> After adjustment for the characteristics of individual patients, hospitals recording the highest readmission rates were those with patients who were the most likely to be readmitted to the hospital due to the complexity of their illness or a low socioeconomic status.<sup>43</sup> In our study, we could not include the socioeconomic status of patients. We are





aware that one plausible explanation for the increase in hospital readmissions could be related to the patient's socioeconomic environment, as social and economic support at home may not be sustained and place the patient at a higher risk of readmission. Indeed, the use of readmission as a marker of complications after an initial surgical admission remains controversial. Some studies reported that almost half of readmissions were not associated with a currently assessed complication.<sup>44</sup> Moreover, readmissions after surgery may be associated with new postdischarge complications related to the procedure rather than exacerbation of complications related to a prior index hospitalisation<sup>45</sup> or confounding issues such as substance abuse or homelessness. Some authors believe that reduced readmission rates alone cannot be used as an indicator of care quality; their effects must be studied more globally to determine whether such reductions coincide with improved quality of life in patients.<sup>46</sup>

To our knowledge, this study is the first to consider all hospital admissions resulting from all-cause readmissions within 30 days over such a long period in a given country. This study nevertheless has certain limitations. First, the global nature of readmission, chosen as an indicator in this study, can only be regarded as a partial assessment of the quality of surgical care. Other measurements should be considered, such as the mortality rate after hospitalisation. Among the readmissions identified, certain were scheduled and did not result from a complication following the first admission. It was not possible to distinguish between scheduled and unscheduled readmissions, because this information is not recorded in French claims data. This is why we decided to exclude admissions for ocular surgery in the M1 model so as to rule out most scheduled readmissions. Second, we could not compute a combined comorbidity score, as suggested by Mehta *et al.*,<sup>47</sup> from the information available in discharge abstracts. Further research is needed, first to characterise readmissions, second to study the influence of the type or the location of hospitals in greater detail,<sup>48</sup> to consider readmissions after outpatient surgery, and finally to better explain the relationship between readmissions and length of hospital stay.<sup>49</sup>

## CONCLUSION

Our nationwide observational study is the first to consider all hospital admissions resulting from all-cause readmissions within 30 days after surgery over such a long period. It suggests that despite the slight temporary rise in readmissions during the implementation of the case-based payment system, this pricing reform does not appear to have had a significant long-lasting effect on readmissions at 30 days in the public sector. The increase in the readmission rate at 30 days after an admission for surgery appears to be related mainly to modifications in care practices, notably for cataract surgery and, second, to a structural modification associated with the ageing patient population. To interpret these results, further studies are

needed to examine the influence of the different changes in populations and clinical practices on readmissions for each group of diseases.

## AVAILABILITY OF DATA AND MATERIALS

The PMSI database was transmitted by the national agency for the management of hospitalisation data (ATH number 2015-111111-47-33). The use of these data by our department was approved by the National Committee for Data Protection. We are not allowed to transmit these data.

## Author affiliations

<sup>1</sup>Department of Biostatistics and Bioinformatics (DIM), University Hospital, Dijon, France

<sup>2</sup>Division of Research and Statistics, Ministry of Health, Paris, France

<sup>3</sup>School of Economics, University of Sorbonne, Paris, France

<sup>4</sup>The Robert F. Wagner School of Public Service, New York University, New York, USA

<sup>5</sup>Laboratoire d'Economie de Dijon, Université Bourgogne-Franche-Comté, Inserm U1200, CNRS UMR 6307, Dijon, France

<sup>6</sup>Department of Thoracic Surgery, University Hospital, Dijon, France

<sup>7</sup>Department of Ophthalmology, University Hospital, Dijon, France

<sup>8</sup>Eye and Nutrition Research Group, Bourgogne Franche-Comté University, Dijon, France

<sup>9</sup>Clinical Investigation Center, Dijon University Hospital, Dijon, France

<sup>10</sup>Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases (B2PHI), INSERM, UVSQ, Institut Pasteur, Université Paris-Saclay, Paris, France

**Contributors** AV and EY conceptualised and designed the study, interpreted the data and wrote the paper. AR contributed substantially to writing the manuscript. VR, MG, AB and CC-G participated in the interpretation of the results and reviewed and revised the manuscript drafts. CQ oversaw the data analysis and interpretation and contributed substantially to writing the manuscript. All authors accept responsibility for the paper as published.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** This study was approved by the National Committee for Data Protection (registration numbers: (a) 1576793; (b) 913291 for Dijon University Hospital and (c) 723116 for the Ministry of Health) and therefore was conducted in accordance with the Declaration of Helsinki. Written consent was not needed for this study.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

1. Fetter RB, Freeman JL. Diagnosis related groups: product line management within hospitals. *Acad Manage Rev* 1986;11:41–54.
2. Fetter RB. Diagnosis Related Groups: Understanding Hospital Performance. *Interfaces* 1991;21:6–26.
3. McMahon LF, Fetter RB, Freeman JL, *et al.* and DRG-based prospective payment. *Hosp Health Serv Adm* 1986;31:62–74.
4. Roger France FH, Roger FH. Case mix use in 25 countries: a migration success but international comparisons failure. *Int J Med Inform* 2003;70(2-3):215–9.



5. Dismuke CE, Sena V. Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods. *Health Care Manag Sci* 1999;2:107–16.
6. Dismuke CE, Guimaraes P. Has the caveat of case-mix based payment influenced the quality of inpatient hospital care in Portugal? *Appl Econ* 2002;34:1301–7.
7. Or Z. Implementation of DRG Payment in France: issues and recent developments. *Health Policy* 2014;117:146–50.
8. Noord P, Hagen T, Iversen T. The Norwegian Health Care System [Internet]. *OECD Economics Department Working Papers* 1998;198:32.
9. Pirson M, Van den Bulcke J, Di Pierdomenico L, et al. Medical and economic evaluation of oncological inpatients in 14 Belgian hospitals]. *Bull Cancer (Paris)*. nov 2015;102:923–31.
10. Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;335:648–50.
11. Thomas JW, Guire KE, Horvat GG. Is patient length of stay related to quality of care? *Hosp Health Serv Adm* 1997;42:489–507.
12. Martin S, Street A, Han L, et al. Have hospital readmissions increased in the face of reductions in length of stay? Evidence from England. *Health Policy* 2016;120:89–99.
13. Kahn KL, Keeler EB, Sherwood MJ, et al. Comparing outcomes of care before and after implementation of the DRG-based prospective payment system. *JAMA* 1990;264:1984–8.
14. Cutler DM. *The Incidence of Adverse Medical Outcomes Under Prospective Payments* [Internet: National Bureau of Economic Research, 1993. Report No.: 4300. <http://www.nber.org/papers/w4300> (accessed 1 Feb 2017).
15. Bahrami S, Holstein J, Chatellier G, et al. Using administrative data to assess the impact of length of stay on readmissions: study of two procedures in surgery and obstetrics]. *Rev Epidemiol Sante Publique*. apr 2008;56:79–85.
16. Moghavern N, McDonald K, Ratliff JK, et al. Performance Measures in Neurosurgical Patient Care: Differing Applications of Patient Safety Indicators. *Med Care* 2016;54:359–64.
17. Rosen AK, Chen Q, Shwartz M, et al. Does Use of a Hospital-wide Readmission Measure Versus Condition-specific Readmission Measures Make a Difference for Hospital Profiling and Payment Penalties? *Med Care* 2016;54:155–61.
18. Serdén L, Lindqvist R, Rosén M. Have DRG-based prospective payment systems influenced the number of secondary diagnoses in health care administrative data? *Health Policy* 2003;65:101–7.
19. Böcking W, Ahrens U, Kirch W, et al. First results of the introduction of DRGs in Germany and overview of experience from other DRG countries. *J Public Health* 2005;13:128–37.
20. Herwartz H, Strumann C. On the effect of prospective payment on local hospital competition in Germany. *Health Care Manag Sci* 2012;15:48–62.
21. Guccio C, Lisi D, Pignataro G. Readmission and Hospital Quality under Prospective Payment System [Internet]. 2014. <https://mpira.uni-muenchen.de/56490/>
22. Abdul-Aziz AA, Hayward RA, Aaronson KD, et al. Association Between Medicare Hospital Readmission Penalties and 30-Day Combined Excess Readmission and Mortality. *JAMA Cardiol* 2017;2:200–3.
23. Desai NR, Ross JS, Kwon JY, et al. Association Between Hospital Penalty Status Under the Hospital Readmission Reduction Program and Readmission Rates for Target and Nontarget Conditions. *JAMA* 2016;316:2647–56.
24. Danino JF, Taylor T, Metcalfe CW, et al. Readmission rates and financial penalties after ear, nose and throat surgery: how can we improve? *Br J Hosp Med* 2015;76:655–7.
25. Gusmano M, Rodwin V, Weisz D, et al. Comparison of rehospitalization rates in France and the United States. *J Health Serv Res Policy* 2015;20:18–25.
26. Lainay C, Benzenine E, Durier J, et al. Hospitalization within the first year after stroke: the Dijon stroke registry. *Stroke* 2015;46:190–6.
27. Delmas MC, Marguet C, Raheison C, et al. Readmissions for asthma in France in 2002–2005. *Rev Mal Respir* 2011;28:e115–e122.
28. Abdulmalak C, Cottenet J, Beltramo G, et al. Haemoptysis in adults: a 5-year study using the French nationwide hospital administrative database. *Eur Respir J* 2015;46:503–11.
29. Creuzot-Garcher C, Benzenine E, Mariet AS, et al. Incidence of Acute Postoperative Endophthalmitis after Cataract Surgery: A Nationwide Study in France from 2005 to 2014. *Ophthalmology* 2016;123:1414–20.
30. Lorgis L, Cottenet J, Molins G, et al. Outcomes after acute myocardial infarction in HIV-infected patients: analysis of data from a French nationwide hospital medical information database. *Circulation* 2013;127:1767–74.
31. Pagès PB, Cottenet J, Mariet AS, et al. In-hospital mortality following lung cancer resection: nationwide administrative database. *Eur Respir J* 2016;47:1809–17.
32. Han S, Smith TS, Gunnar W. Descriptive analysis of 30-day readmission after inpatient surgery discharge in the Veterans Health Administration. *JAMA Surg* 2014;149:1162–8.
33. Davis C, Rhodes DJ. The impact of DRGs on the cost and quality of health care in the United States. *Health Policy* 1988;9:117–31.
34. DesHarnais S, Kobrinski E, Chesney J, et al. The early effects of the prospective payment system on inpatient utilization and the quality of care. *Inquiry* 1987;24:7–16.
35. Sager MA, Easterling DV, Kindig DA, et al. Changes in the location of death after passage of Medicare's prospective payment system. A national study. *N Engl J Med* 1989;320:433–9.
36. O'Reilly J, Busse R, Häkkinen U, et al. Paying for hospital care: the experience with implementing activity-based funding in five European countries. *Health Econ Policy Law* 2012;7:73–101.
37. Reuter PG, Kernéis S, Turbelin C, et al. [Orientation of patients referred by their general practitioner to the public or private hospital sector in France: A prospective epidemiologic study]. *Rev Med Interne* 2012;33:672–7.
38. DeVore AD, Hammill BG, Hardy NC, et al. Has Public Reporting of Hospital Readmission Rates Affected Patient Outcomes?: Analysis of Medicare Claims Data. *J Am Coll Cardiol* 2016;67:963–72.
39. McGarry BE, Blankley AA, Li Y. The Impact of the Medicare Hospital Readmission Reduction Program in New York State. *Med Care* 2016;54:162–71.
40. Hofstede SN, van Bodegom-Vos L, Kringos DS, et al. Mortality, readmission and length of stay have different relationships using hospital-level versus patient-level data: an example of the ecological fallacy affecting hospital performance indicators. *BMJ Qual Saf* 2017;doi: 10.1136/bmjqs-2017-006776. [Epub ahead of print].
41. Lu N, Huang KC, Johnson JA. Reducing excess readmissions: promising effect of hospital readmissions reduction program in US hospitals. *Int J Qual Health Care* 2016;28:53–8.
42. Barnett ML, Hsu J, McWilliams JM. Patient characteristics and differences in hospital readmission rates. *JAMA Intern Med* 2015;175:1803–12.
43. Leape LL. Hospital readmissions following surgery: turning complications into "treasures". *JAMA* 2015;313:467–8.
44. Morris MS, Deierhoi RJ, Richman JS, et al. The relationship between timing of surgical complications and hospital readmission. *JAMA Surg* 2014;149:348–54.
45. McIntyre LK, Arbabi S, Robinson EF, et al. Analysis of risk factors for patient readmission 30 days following discharge from general surgery. *JAMA Surg* 2016;151:855–61.
46. Burgess JF, Hockenberry JM. Can all cause readmission policy improve quality or lower expenditures? A historical perspective on current initiatives. *Health Econ Policy Law* 2014;9:193–213.
47. Mehta HB, Dimou F, Adhikari D, et al. Comparison of comorbidity scores in predicting surgical outcomes. *Med Care* 2016;54:180–7.
48. Metcalfe D, Olufajo OA, Zogg CK, et al. Are older adults with hip fractures disadvantaged in level 1 trauma centers? *Med Care* 2016;54:616–22.
49. Sedrakyan A, Kamel H, Mao J, et al. Hospital readmission and length of stay over time in patients undergoing major cardiovascular and orthopedic surgery: a tale of 2 states. *Med Care* 2016;54:592–9.

## **Devenir à un an des patients admis pour tentative de suicide**

Article publié dans *Epidemiology and psychiatric sciences*, 2019

### Résumé

#### Objectif

Les lésions auto-infligées sont un motif fréquent de recours en urgence à une prise en charge hospitalière. L'analyse de la fréquence des récidives et de la mortalité dans la période qui suit la sortie de l'hospitalisation initiale en court séjour, à partir d'un recueil systématique à large échelle, est susceptible d'apporter des informations pouvant guider la pratique clinique et éclairer l'organisation plus générale de la prise en charge de ces situations.

#### Méthode

Tous les séjours avec un code de lésion auto-infligée (X60 à X84 en CIM-10) figurant dans une base de données correspondant à 70% de l'activité hospitalière de court séjour en France pour les années 2008 et 2009, ont été sélectionnés. Cette base de données (Amphi), constituée pour l'analyse de la mortalité post-hospitalière, est issue de l'appariement de la base de données PMSI nationale pour ces années et ce secteur à la base de données de l'Assurance maladie qui comporte la date d'un éventuel décès pour les personnes affiliées au régime général (70% de la population française), et dans ce cas, un appariement à la base de données des causes médicales de décès a pu être réalisé dans 96% des cas. La base de données Amphi donne une l'information sur le statu vital dans l'année qui suit chaque hospitalisation. Le premier séjour de chaque patient est considéré comme le séjour index et est retenu dans la mesure où le patient est sorti vivant.

#### Résultat

136451 de tels séjours figurent dans la base Amphi. Les caractéristiques des patients sont les suivantes : 62,8% de femmes, un âge médian de 38 ans chez les hommes comme les femmes ; pour les femmes la distribution des âges est marquée par 2 pics, à 16 et 44 ans ; le moyen principal est l'intoxication médicamenteuse volontaire (82.1%) ; 12,9% des patients ont été admis en réanimation. Au décours de cette hospitalisation, 71,3% des patients sont retournés à domicile, 23,7% ont été transférés en psychiatrie. Dans l'année qui a suivi, 12,4% des patients ont été réadmis pour lésion auto-infligée, dont 75,2% dans les 6 premiers mois, et 74,6% avec une seule réadmission sur l'année de suivi. La mortalité globale dans l'année a été de 2,6% pour un ratio standardisé de mortalité de 7,5 (20 pour les jeunes adultes). Les causes de décès se répartissent de la façon suivante : maladie 35,7% suicide 34,4%, inconnue 17,5%, traumatique 12,4%. 62,9% des décès par suicide surviennent dans les 6 premiers mois. Les décès par suicide correspondent à l'utilisation d'un moyen violent (i.e. autre que l'intoxication médicamenteuse volontaire) pour 70%, la concordance entre moyen utilisé lors du séjour index et celui du suicide est de 30%. Les facteurs de risque pour un décès par suicide sont l'âge, le sexe masculin, l'utilisation première d'un moyen violent, l'admission en réanimation, une sortie par transfert. La valeur prédictive positive du modèle est cependant faible, la valeur discriminante du modèle n'étant pas suffisante (AUC-ROC 0,77) pour des événements de fréquence faible.

#### Conclusion

Si le risque de décès par suicide est faible pour les patients à l'issue d'une hospitalisation pour lésion auto-infligée, il est cependant beaucoup plus élevé que dans la population générale, indiquant la vulnérabilité de ces personnes. Le risque apparait le plus élevé dans les 6 premiers mois, ce qui engage à la mise en place d'actions de prévention sur cette période critique.



## Recurrence and mortality 1 year after hospital admission for non-fatal self-harm: a nationwide population-based study

A. Vuagnat<sup>1,2,\*</sup>, F. Jollant<sup>3,4,\*</sup>, M. Abbar<sup>5</sup>, K. Hawton<sup>6</sup> and C. Quantin<sup>2,7,8,9</sup>

### Original Article

\*These authors equally contributed to this work

**Cite this article:** Vuagnat A, Jollant F, Abbar M, Hawton K, Quantin C (2019). Recurrence and mortality 1 year after hospital admission for non-fatal self-harm: a nationwide population-based study. *Epidemiology and Psychiatric Sciences* 1–10. <https://doi.org/10.1017/S2045796019000039>

Received: 6 August 2018

Revised: 17 January 2019

Accepted: 17 January 2019

#### Key words:

mortality; recurrence; self-harm; suicide

#### Author for correspondence:

F. Jollant, E-mail: [fabrice.jollant@parisdescartes.fr](mailto:fabrice.jollant@parisdescartes.fr)

<sup>1</sup>Department for Research, Studies, Evaluation and Statistics (DREES), French Health and Social Affairs Ministry, Paris, France; <sup>2</sup>Biostatistics and Bioinformatics (DIM), University Hospital, Dijon & France Bourgogne Franche-Comté University, Dijon, France; <sup>3</sup>Paris-Descartes University, & Sainte-Anne Hospital, Paris, France; <sup>4</sup>McGill group for suicide studies, McGill University, Montréal, Canada; <sup>5</sup>Academic Hospital (CHU) Nîmes, Nîmes, France; <sup>6</sup>Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK; <sup>7</sup>Inserm, CIC 1432, Dijon, France; <sup>8</sup>Dijon University Hospital, Clinical Investigation Center, clinical epidemiology/clinical trials unit, Dijon, France and <sup>9</sup>Biostatistics, Biomathematics, Pharmaco-epidemiology and Infectious Diseases (B2PHI), Inserm, UVSQ, Institut Pasteur, Université Paris-Saclay, Paris, France

#### Abstract

**Aims.** A large number of people present each day at hospitals for non-fatal deliberate self-harm (DSH). Examination of the short-term risk of non-fatal recurrence and mortality at the national level is of major importance for both individual medical decision-making and global organisation of care.

**Methods.** Following the almost exhaustive linkage (96%) of two national registries in France covering 45 million inhabitants (i.e. 70% of the whole population), information about hospitalisation for DSH in 2008–2009 and vital status at 1 year was obtained. Individuals who died during the index hospital stay were excluded from analyses.

**Results.** Over 2 years, 136,451 individuals were hospitalised in medicine or surgery for DSH. The sample comprised 62.8% women, median age 38 in both genders, with two peaks at 16 and 44 years in women, and one peak at 37 years in men. The method used for DSH was drug overdose in 82.1% of cases. Admission to an intensive care unit occurred in 12.9%. Following index hospitalisation, 71.3% returned home and 23.7% were transferred to a psychiatric inpatient care unit. DSH recurrence during the following year occurred in 12.4% of the sample, within the first 6 months in 75.2%, and only once in 74.6%. At 1 year, 2.6% of the sample had died. The overall standardised mortality ratio was 7.5 but reached more than 20 in young adults. The causes were natural causes (35.7%), suicide (34.4%), unspecified cause (17.5%) and accident (12.4%). Most (62.9%) deaths by suicide occurred within the first 6 months following index DSH. Violent means (i.e. not drug overdose) were used in 70% of suicide cases. Concordance between means used for index DSH and for suicide was low (30% overall), except for drug overdose. Main suicide risk factors were older age, being male, use of a violent means at index DSH, index admission to an intensive care unit, a transfer to another medical department or to a psychiatric inpatient unit, and recurrence of DSH. However, these factors had low positive predictive values individually (below 2%).

**Conclusions.** Non-fatal DSH represent frequent events with a significant risk of short-term recurrence and death from various causes. The first 6 months following hospital discharge appear to be a critical period. Specific short-term aftercare programs targeting all people with a DSH episode have to be developed, along other suicide prevention strategies.

#### Introduction

Non-fatal deliberate self-harm (DSH) represents considerable public health issues. The lifetime prevalence of DSH has been estimated at 1.3% of the general population in Europe (Bernal *et al.*, 2007). In addition, people who present to hospital for non-fatal DSH are at increased risk of premature death. Studies have robustly shown that these acts are the strongest risk factor for future suicide (Cavanagh *et al.*, 2003). A recent meta-analysis based on 40 studies estimated the risk of subsequent suicide death during the following year around 1.6% (CI 1.2–2.4) (Carroll *et al.*, 2014). Rates of suicide in those who self-harmed during the previous year are therefore in the range of 100 times the rates in the general population in Europe and North America (Owens *et al.*, 2002). However, excess mortality in people who self-harmed is not solely attributable to suicide but also to natural death (Hawton *et al.*, 2006). People who self-harm, therefore, constitute a particular fragile population needing special care. Overall, the Global Burden of Diseases study estimated that DSH ranks among the ten leading causes of disability-adjusted life years (DALYs) in people aged 15–39 worldwide (GBD\_2015, 2016). The annual economic cost of DSH is high, reaching for instance \$93 billion in total in the

USA (Shepard *et al.*, 2016), and 162 million pounds in England for general hospital costs only (Tsiachristas *et al.*, 2017). Thus, an accurate estimation of the phenomenon is particularly important to assist organisation of care and prevention.

In the present study, we aimed to study mortality following DSH at a national level in order to limit potential biases related to local conditions. Indeed, while most previous studies were based on local data, very few used national registries (Payne *et al.*, 2009; Runeson *et al.*, 2010; Chung *et al.*, 2012; Perry *et al.*, 2012). To our knowledge, the largest cohort studies comprised around 40 000–60 000 admissions or individuals in Scotland, Sweden, Ireland and Denmark (Payne *et al.*, 2009; Runeson *et al.*, 2010; Perry *et al.*, 2012; Fedyszyn *et al.*, 2016), and in Medicaid-financed adults in the USA (Olson *et al.*, 2017). Our database includes exhaustive information about approximately 45 million inhabitants in France covered by the general program of the public insurance system – i.e. 70% of the French population (the rest of the population being covered by specific insurance programs, whose data are currently not accessible) – over 2 consecutive years. The strength of this retrospective cohort is the very limited number of cases for whom the vital status is missing. The size of the sample allowed us to (i) examine DSH at a national scale in a country with an excellent universal insurance system; (ii) investigate the risk of short-term non-fatal DSH repetition; (iii) measure mortality rates at 1 year, in relation to suicide, accidents and natural causes; and (iv) weight risk factors including rare events like particular types of suicidal acts (drowning, gas, etc.). Regarding the high number of patients hospitalised for DSH each year in Western countries, we believe this study provides valuable information for clinicians working in Emergency departments, intensive care units or in liaison-consultation psychiatry for medical decision-making and to health institutions for care organization.

## Methods

### Population and database linkage

Data were extracted from two national registries. Medico-administrative data on all individuals hospitalised in medicine and surgery departments in France are routinely collected in a national database named 'Programme de Médicalisation des Systèmes d'Information' (PMSI). This database collects information about individuals' name, date of birth, gender, place of residence, hospital location and diagnoses based on International Classification of Diseases–10<sup>th</sup> edition (ICD-10) codes. A second database named 'CepiDc' monitored by the National Institute for Health Research (Inserm) provides individual information about date and cause of death, in addition to socio-demographic data (name, gender, date of birth) and geographical variables (place of death and place of residence). This information is based on death certificates completed by physicians and forensic institutes. The linkage of these two exhaustive databases is to date only possible for individuals under the general program of the national health insurance system, i.e. approximately 70% of the French population (45 million people in 2009). The process resulted from a recent collaboration between different administrations under a project named 'Amphi' in order to study post-hospitalisation mortality. This study was approved by the National Committee for data protection (registration number: 1454315). Patient written consent was not necessary for this study.

We extracted data on individuals hospitalised for non-fatal DSH – corresponding to ICD-10 codes X60 to X84 – from 1

January 2008 to 31 December 2009. This extraction yielded 168 053 hospitalisations after exclusion of patients who died during the index hospitalisation, corresponding to 136 451 individuals and 0.7% of the total number of hospitalisations for the study population during this period. Based on probabilistic linkage according to different variables (the month and year of birth, full date of death, gender and residency location), mortality data could be obtained from CepiDc for 96% of the sample (see flow chart in Fig. 1). In order to investigate the representativeness of the data obtained from the Amphi database used in the current study in comparison with the whole PMSI database, we compared age and gender distributions in people who presented with DSH and found a very satisfactory overlap (data available on request).

### Variables

We considered the first hospitalisation for DSH within the studied period as the index DSH episode. For each patient, socio-demographic and medical data were collected at the index hospitalisation: age, gender, type of hospital, distance from home to hospital, type of living place according to the zip-code, admission to intensive care unit, Charlson comorbidity index (Charlson *et al.*, 1987), duration of hospital stay, transfer following initial care and means used for the index DSH episode. In the dataset, there were no missing data for all these variables except for zip-code (0.4% of missing data). Psychiatric diagnoses were not retained due to a high number of data missing (53% of individuals) and an expected lack of reliability in the way diagnoses were carried out in Emergency departments and medical/surgery facilities.

Some DSH means were pooled to facilitate analyses. The following categories were retained: intentional self-poisoning by drugs (X60 to 64), alcohol (X65), an unspecified chemicals or noxious substances (X66 and X68 to 69), gases and vapours (X67); intentional self-harm by hanging, strangulation or suffocation (X70), drowning and submersion (X71), firearm discharge (X72 to 74), use of explosive material, smoke, fire and flames, and steam, hot vapours and hot objects (X75 to 77), cutting with sharp or blunt objects (X78–79), jumping from a high place (X80), jumping or lying in front of moving objects or crashing of motor vehicle (X81–82), other specified means (X83) and unspecified means (X84). Violent means were all means except drugs and alcohol (X66 to 82).

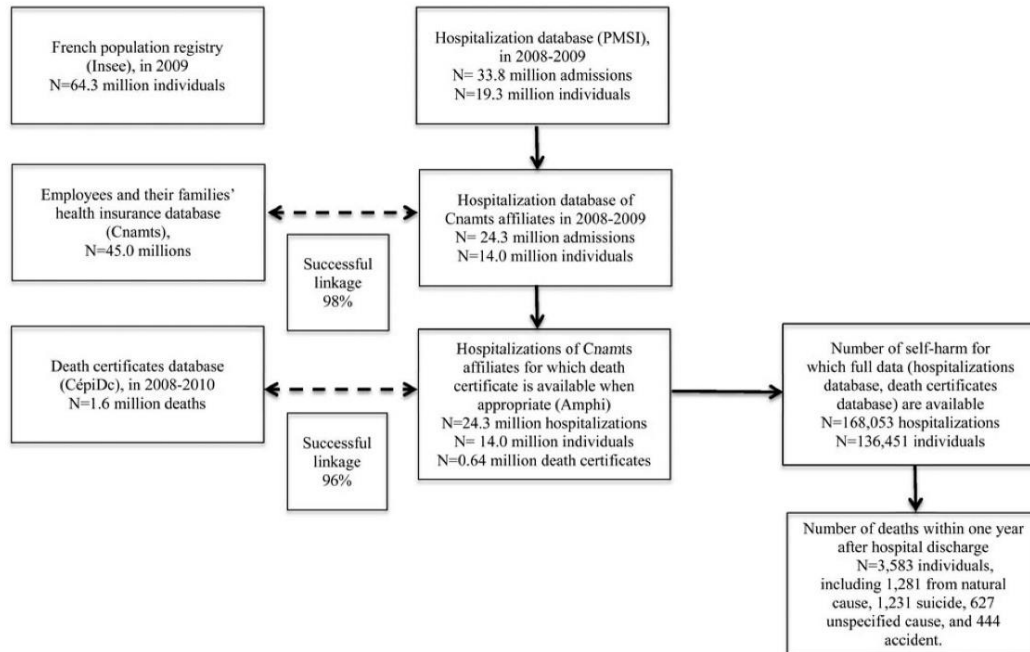
Mortality and DSH recurrence were then examined within a 1-year period following the index DSH episode for each individual. Mortality data were collected for the period 2008–2010. Missing information on patient death 1 year after hospital discharge is only 0.1%. Causes of deaths were classified as suicide, accident, natural causes and undetermined causes.

### Statistical analyses

We present categorical variables as frequency distributions and continuous variables as means (with standard deviations, SD) or medians.

We developed logistic regression models to analyse the impact of factors associated with overall mortality and the four main causes of death, based on univariate analyses. In multivariate analyses, we introduced all individual variables considered significant in univariate analyses ( $p < 0.20$ ) according to their clinical relevance based on correlation tests. We used backward selection.





**Fig. 1.** Flow chart representing the different steps of data collection.

*Footnotes:* Insee: Institut National de la Statistique et des Etudes Economiques; Cnamts: Caisse Nationale d'Assurance Maladie des Travailleurs Salariés; CepiDc: Centre d'Epidémiologie sur les causes médicales de décès; PMSI: Programme de Médicalisation des Systèmes d'Information; Amphi: Amphi database.

The overall accuracy of the model was assessed by the *c*-statistic, which represents the area under the curve (ROC). Finally, we analysed the impact of factors associated with a new hospitalisation for a recurrence of a DSH occurring within 1 year after the index hospitalisation. In order to analyse the potential time-dependent effect of recurrence of a DSH, we also used a Cox regression model including the same variables as above.

Proportional hazards assumption was checked for each variable of the Cox model, using graphs plotting the log of the negative log of the Kaplan-Meier survival function estimates *v.* log of the time to event, as well as including an interaction with log of the time to event in the model, considered as a sensitivity analysis. The proportional hazard assumption was satisfied for nearly all variables except for the duration of hospital stay. After taking into account this interaction in the final model, the hazard ratio estimates were almost identical to the results presented below. Only the effect of the duration of hospital stay was modified as longer durations tended to be associated with an increasing risk over time.

Although psychiatric diagnoses were not included as explicative variables in the main analyses for reasons exposed above, we performed a second sensitivity analysis with these variables in order to test the robustness of the model. This sensitivity analysis showed no significant change in the estimation of hazard ratios of the studied variables.

Finally, a Standardised Mortality Ratio (SMR) was calculated for each age group and gender as the ratio of death observed in the sample *v.* expected deaths in the general population.

## Results

### Description of the sample at index hospital admission

Description of the sample at the time of index admission can be found in Table 1 and distribution of hospital admissions for DSH is shown in Fig. 2.

### Recurrence of DSH within 1 year following index DSH

A majority of individuals were not re-hospitalised for a new DSH within the following year as recurrence was found in only 12.4% ( $N = 16\,919$ ) of the sample. Median delay between the index act and the following one was 84 days. The first recurrence happened within the first week in 11.4% of people who re-presented, within the first month in 28.0%, within the first 3 months in 52.4%, and within the first 6 months in 75.2%. Most individuals used the same means between the index act and the next DSH (78.1%), while 6.6% switched from a non-violent to a violent means and 5.1% from a violent to a non-violent means (unspecified in 10.2%).

Most DSH recurrence occurred only once within the 1-year follow-up ( $N = 12\,616$ , 74.6%) while 16.1% re-presented a second time, and 9.3% more than 3 times, including 2.1% more than 5 times.

### One-year mortality following DSH

The death occurred in 2.6% ( $N = 3\,583$ ) of the total sample within 1 year following the index DSH episode. Figure 3 shows the global



**Table 1.** Description of the sample at the time of index admission

Variables	N = 136 451
Gender, female (%)	62.8
Age, median [min-max]	38 [9–94]
Age distribution (%):	
≤25	26.0
26 to 64	67.9
≥65	6.1
Type of living place (%):	
City centre	48.0
Suburbs	32.8
Small towns	10.6
Rural areas	8.6
Means used for index DSH (%)*:	
Drugs	82.1
Unspecified means	5.6
Alcohol ingestion	5.0
Cutting with sharp or blunt objects	4.9
Ingestion of unspecified chemicals or noxious substances	2.4
Hanging, strangulation or suffocation	1.6
Jumping from a high place	0.9
Firearm discharge	0.4
Other specified means	0.4
Drowning and submersion	0.3
Gases and vapours	0.3
Jumping or lying in front of moving objects or crashing of motor vehicles	0.2
Use of explosive material, smoke, fire and flames, and steam, hot vapours and hot objects	0.2
Use of a violent means (%)	11.2
Only means codes (%)	91.7
Type of index hospitalisation (%):	
General hospital	71.9
Academic hospital (CHU)	23.7
Private clinic	4.4
Admission to an intensive care unit (%)	12.9
Charlson Comorbidity index ≥0	6.0
Mean Charlson index score when different from 0 (sd)	2.0 (1.4)
Duration of index hospital stay (%):	
<24 h	16.8
1 days	51.0
2–7 days	25.3
>7 days	6.9
Mean duration of hospital stay (sd)	2.4 (4.1)

(Continued)

**Table 1.** (Continued.)

Variables	N = 136 451
Following hospitalisation (%):	
Discharged home	71.3
Transfer to psychiatric hospital in-patient care	23.7
Transfer to another medical or surgery department	5.0

Footnotes: \*Some individuals are counted twice in case of use of multiple means.

survival curve over the 1-year follow-up. The SMR relative to expected deaths in the general population was 8.5 in men (from 3.3 in those aged 76 and above, to 24.5 in 26–35 s), 6.6 in women (from 2.8 in 76+ to 28.9 in 26–35 s) and 7.5 overall.

Causes of deaths were: natural causes (35.7% of death cases), suicide (34.4%), unspecified (17.5%) and accident (12.4%). Among men, 3.9% had died within 1 year, including 1.4% due to suicide and 1.3% due to natural causes. The equivalent figures for women were 1.9, 0.6 and 0.7%, respectively. Figure 1 in supplemental material represents the number of deaths per age group and gender for each cause.

The means used for suicide were hanging, strangulation or suffocation in 42.1%, drugs in 28.4%, jumping from a height in 6.3%, drowning and submersion 5.4%, firearm discharge 5.0%, ingestion of unspecified chemicals or noxious substances 2.5%, jumping or lying in front of moving objects or crashing of motor vehicle 1.7%, cutting with sharp or blunt objects 1.5%, use of explosive material, smoke, fire and flames, steam 1.1%, self-poisoning by gases and vapours 0.7%, ingestion of alcohol 0.1%, other specified means 0.3% and unspecified means 4.9%.

Overall concordance of suicidal means between the index DSH and suicide was 30.2% but with large differences according to the means used: 79.9% of those who died by suicide using drugs previously used the same methods for the index DSH, while the proportion was 35.5% for chemicals, 22.6% for firearm, 15.8% for cutting with sharp or blunt objects, 9.8% for hanging and below 8% for all other means. It is important to note that the majority of individuals who died by suicide had used drug overdose at the index DSH (N = 878 among 1231 suicide deaths, 71.3%).

Death by suicide occurred within the first week following the index DSH in 5.5%, within the first month in 20.5%, within the first 3 months in 41.7%, and within the first 6 months in 62.9%. The cumulative risk for the other causes of death showed a more linear distribution over the 1-year follow-up (Fig. 2 in Supplemental material).

### Predictive factors for 1-year mortality

Tables 2 and 3 present the results of logistic regression analyses for overall mortality and the four main causes of death. While violent suicidal means at the index DSH index and admission to an intensive care unit were both significantly associated with future suicide, their positive predictive values were low (1.6% and 1.6%, respectively). However, the global model including all factors of the regression analysis yielded a fair-good accuracy with an area under the curve of 0.77.

Cox regression model showed that recurrence of a DSH was also associated with an increased risk of overall mortality (HR = 1.98 [1.78–2.21]) and was related to suicide (HR = 2.30 [1.92–2.75]) but also accident (HR = 2.70 [2.06–3.55]), natural

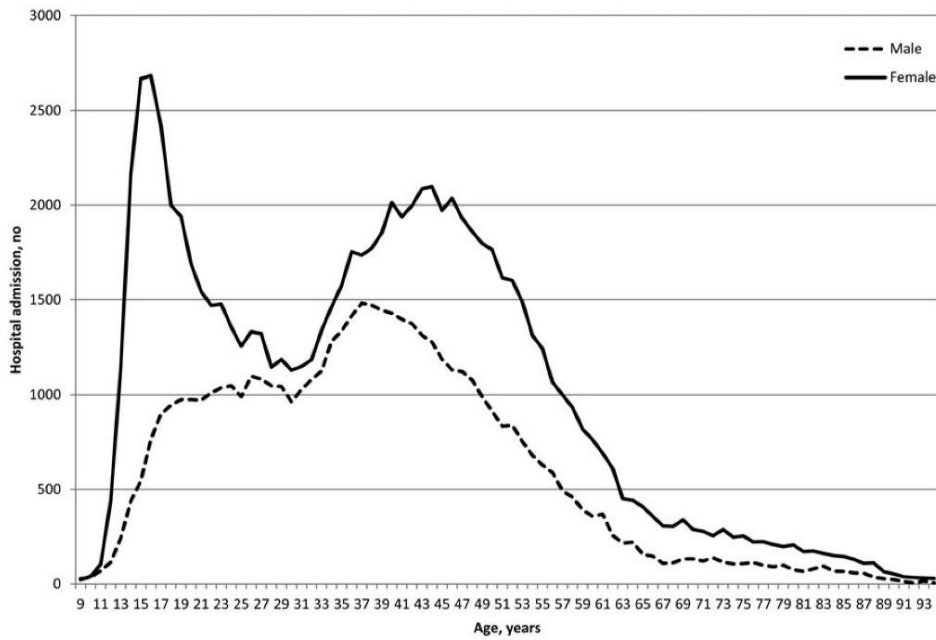


Fig. 2. Distribution of age at index hospital admission for deliberate self-harm, by gender.

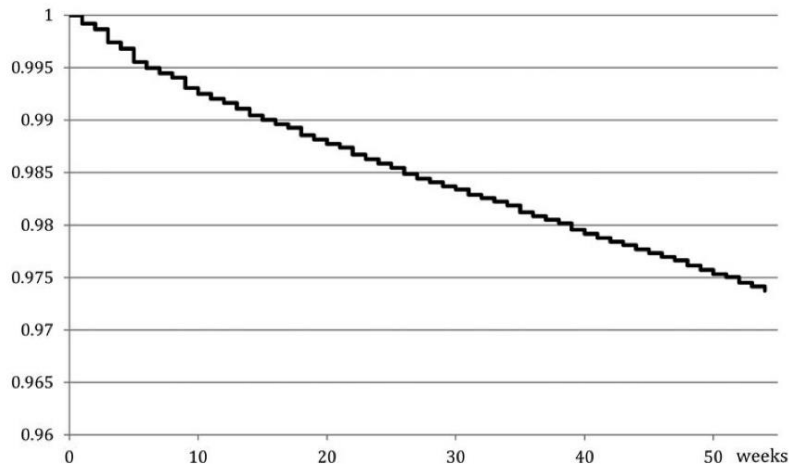


Fig. 3. Global survival curve since index hospital admission for deliberate self-harm.

causes (HR = 1.44 [1.15–1.79]) and unspecified causes (HR = 2.15 [1.68–2.75]).

Detailed results according to DSH means are presented in Supplemental material.

**Discussion**

The present study is the largest one published to date in terms of a number of individuals followed over 1 year after a DSH and for

whom mortality data are available. Other strengths of this sample are the extent of the national coverage (almost 70% of the French population), the totally exhaustive data collection of hospitalisations and the very limited number of cases for which vital status was not known after 1 year (4%). Main results of this study are: (1) At index admission, individuals who self-harmed were mainly women (63%) with a median age of 38 and a narrow peak at 16, using medication overdose (82.1%). Most (71.3%) returned home the same day or after a short stay at the hospital (1 day or less for

**Table 2.** Risk factors for overall mortality and suicide

Variables	All causes				Suicide			
	Hazard ratio	95% CI		<i>p</i>	Hazard ratio	95% CI		<i>p</i>
Age (reference = 9–15 year-old)								
16–25	4.0	2.0	7.9	<0.0001	3.6	1.5	9.0	0.005
26–35	11.9	6.1	23.1	<0.0001	8.3	3.4	20.4	<0.0001
36–45	16.3	8.4	31.6	<0.0001	12.2	5.0	29.5	<0.0001
46–55	22.9	11.9	44.2	<0.0001	17.0	7.0	41.2	<0.0001
56–65	32.6	16.8	63.0	<0.0001	19.0	7.8	46.3	<0.0001
66–75	49.8	25.7	96.7	<0.0001	28.0	11.3	69.1	<0.0001
76 +	92.0	47.6	178.2	<0.0001	29.3	11.8	72.9	<0.0001
Gender: Male v. female	2.0	1.9	2.1	<0.0001	2.1	1.9	2.4	<0.0001
Means used for index DSH (ref = drugs [X60-X64])								
Violent means [X66-X82]	1.4	1.2	1.5	<0.0001	1.5	1.3	1.8	<0.0001
Alcohol [X65]	1.1	1.0	1.3	0.2	0.4	0.3	0.7	<0.0001
More than one DSH means (v. single)	0.9	0.7	1.0	0.1	1.3	1.0	1.7	0.1
University v. general hospital	1.0	1.0	1.0	0.4	0.8	0.7	0.9	0.002
Private clinic v. general hospital	0.9	0.7	1.0	0.06	0.9	0.7	1.2	0.5
Distance from home to hospital (ref = 0–3 km)								
4–12 km	0.9	0.9	1.0	0.2	1.1	0.9	1.3	0.2
13 km and more	0.9	0.9	1.0	0.2	1.1	0.9	1.2	0.4
Living place (ref = city centre)								
Suburbs	1.0	0.9	1.0	0.4	0.9	0.8	1.0	0.05
Small town	1.0	0.9	1.1	0.9	1.0	0.9	1.3	0.5
Rural	0.8	0.7	1.0	0.01	0.8	0.7	1.0	0.1
Admission to intensive care unit (yes v. no)	1.2	1.1	1.3	<0.0001	1.5	1.3	1.7	<0.0001
Comorbidity (ref : Charlson index = 0)								
Charlson = 1–2	1.6	1.4	1.7	<0.0001	0.8	0.6	1.0	0.02
Charlson = 3–4	1.9	1.6	2.3	<0.0001	0.4	0.2	0.8	0.01
Charlson = 5 and more	5.5	4.6	6.5	<0.0001	1.3	0.7	2.5	0.4
Duration of hospital stay (ref = 0–1 night)								
2–3 nights	1.1	1.0	1.2	0.06	1.2	1.0	1.4	0.03
4 and more nights	1.1	1.0	1.2	0.1	1.1	0.9	1.3	0.4
Transfer following initial care (ref = home return)								
to another medical department	2.0	1.8	2.2	<0.0001	2.3	1.9	2.8	<0.0001
to a psychiatric inpatient care unit	1.2	1.1	1.3	<0.0001	1.7	1.5	1.9	<0.0001

almost 70%). (2) Recurrence of DSH within 1 year occurred in 12.4% of cases, during the first 6 months in 75.2% of cases and usually once only. (3) Mortality at 1 year reached 2.6% of the sample, shared between suicide and other causes. In 63% of cases, suicide occurred within the first 6 months. Classical risk factors were identified but were of extremely limited predictive value.

Several limitations have to be highlighted before discussing findings. First, data for about 30% of the French population were not available, limiting representativeness. This concerns individuals with specific jobs – which currently benefit from special insurance programs – notably farmers, self-employed

workers, most students, or soldiers. While the comparison of the distributions of age and gender in individuals with DSH affiliated or not to the general program did not show significant differences, the modulating effect of specific professional groups on particular variables, especially suicide risk, cannot be excluded. Second, while individuals who engage in DSH acts are usually referred to an emergency department for physical examination and the most severe acts are hospitalised even for a few hours, an unknown number of DSH episodes are not. This includes patients directly referred to a psychiatric hospital and those not hospitalised. Moreover, the identification of cases with DSH solely

**Table 3.** Risk factors for death from natural causes, accident and unspecified cause

Variables	Accident			Natural causes			Unspecified cause					
	Hazard ratio	95% CI	<i>p</i>	Hazard ratio	95% CI	<i>p</i>	Hazard ratio	95% CI	<i>p</i>			
Age (ref = 9–15 year-old)												
16–25	2.8	0.7	11.9	0.2	3.7	0.5	28.6	0.2	6.8	0.9	49.9	0.06
26–35	9.4	2.3	38.4	0.002	15.2	2.1	110.5	0.007	25.4	3.5	182.5	0.001
36–45	9.3	2.3	37.8	0.002	34.0	4.8	243.0	0.0004	30.5	4.3	218.8	0.0007
46–55	8.9	2.2	36.3	0.002	64.7	9.1	461.5	<0.0001	39.7	5.5	284.1	0.0002
56–65	14.2	3.5	58.5	0.0002	127.7	17.9	910.9	<0.0001	46.1	6.4	332.3	0.0001
66–75	13.1	3.1	56.5	0.0005	226.9	31.8	1620.1	<0.0001	58.3	8.0	425.6	<0.0001
76+	42.7	10.3	176.3	<0.0001	449.0	63.0	3199.9	<0.0001	139.0	19.2	1003.1	<0.0001
Gender: Male vs. female	2.2	1.8	2.7	<0.0001	1.829	1.6	2.1	<0.0001	2.0	1.7	2.4	<0.0001
Means used for index DSH (ref = drugs [X60-X64])												
Violent means [X66-X82]	1.2	0.9	1.5	0.3	1.3	1.1	1.5	0.001	1.3	1.0	1.7	0.02
Alcohol [X65]	1.7	1.1	2.5	0.01	1.4	1.2	1.8	0.001	1.3	0.9	1.9	0.1
More than one DSH means (v. single)	0.7	0.4	1.1	0.1	0.8	0.6	1.1	0.2	0.7	0.4	1.1	0.1
University v. general hospital	0.9	0.7	1.1	0.4	0.9	0.8	1.1	0.3	1.5	1.2	1.8	<0.0001
Private clinic v. general hospital	0.4	0.2	0.8	0.006	0.9	0.7	1.2	0.4	0.9	0.6	1.4	0.8
Distance from home to hospital (ref = 0–3 km)												
4–12 km	0.8	0.6	1.1	0.1	1.0	0.8	1.1	0.5	0.8	0.6	0.9	0.009
13 km and more	0.9	0.7	1.2	0.6	0.9	0.8	1.1	0.3	0.8	0.6	1.0	0.03
Living place (ref = city centre)												
Suburbs	1.1	0.9	1.4	0.2	1.0	0.8	1.1	0.6	1.1	0.9	1.3	0.5
Small town	1.0	0.7	1.4	0.9	1.0	0.8	1.2	0.6	0.8	0.6	1.1	0.2
Rural	0.8	0.5	1.2	0.2	1.0	0.7	1.1	0.3	0.7	0.5	1.0	0.04
Admission to intensive care unit (yes v. no)	1.9	1.4	2.4	<0.0001	0.9	0.8	1.1	0.2	1.4	1.1	1.7	0.003
Comorbidity (ref : Charlson index = 0)												
Charlson = 1–2	1.4	1.0	1.9	0.06	2.7	2.4	3.2	<0.0001	1.3	1.0	1.7	0.08
Charlson = 3–4	2.1	1.2	3.7	0.009	3.9	3.1	4.8	<0.0001	0.6	0.3	1.3	0.2
Charlson = 5+	1.6	0.6	4.5	0.3	13.7	11.1	16.9	<0.0001	3.5	2.1	5.8	<0.0001
Duration of hospital stay (ref = 0–1 night)												
2–3 nights	0.8	0.6	1.1	0.2	1.2	1.0	1.5	0.01	0.8	0.7	1.1	0.1
4 and more nights	0.7	0.5	0.9	0.008	1.2	1.0	1.4	0.04	1.0	0.8	1.3	0.7
Transfer following initial care (ref = home return)												
to another medical department	2.4	1.8	3.2	<0.0001	1.9	1.6	2.1	<0.0001	1.6	1.2	2.0	0.001
to a psychiatric inpatient care unit	1.3	1.0	1.6	0.03	0.8	0.7	1.0	0.02	1.1	0.9	1.3	0.2



relies on the coding made by hospital physicians. These factors lead to an underestimation of the DSH phenomenon only based on hospital reports, as previously shown (Geulayov *et al.*, 2018). Third, the misclassification of causes and undetermined causes are major issues in suicide research (Kapusta *et al.*, 2011). For instance, a study in England showed that only 65% of suicide cases as determined by researchers had been given a suicide verdict by coroners, and coroners were less likely to give suicide verdicts in cases of poisoning, drowning and jumping (Palmer *et al.*, 2015). In our study, this could explain a strong association between DSH by drowning and future accidental death but not suicide, or the lack of association between DSH by jumping and future suicide. Fourth, no information is currently available in administrative databases on particular known suicide risk factors, such as suicidal intent, personality traits, a history of childhood abuse, recent life events, among others. Also, some data could not be retained for analyses. This is notably the case for psychiatric diagnoses regarding their expected lack of reliability from emergency units and other medico-surgery facilities. Furthermore, there is a lack of current access to potential outpatient psychiatric care during the 1-year follow-up. Moreover, we had no information about a history of DSH before the index admission. Index admission in the current study is therefore not synonymous as a personal history of first hospitalisation for DSH. Finally, 2008–2009 was the only period for which full data were available with a follow-up of 1 year at the time of analyses. More data will be available in the next years. Of note, recent reductions in DSH and suicide rates in France suggest that prevalence rates are currently lower. Keeping in mind these limitations, our study yielded or confirmed several valuable findings at a large scale.

An important number of individuals were hospitalised for DSH in France in 2008 and 2009. Our sample was largely similar to previous studies conducted in different countries (Hawton *et al.*, 2015) with almost 63% of women, and 50% were aged below 38 years (including 26% below 25 years). There was a notable peak in females aged 16 years. Almost 80% of individuals used self-poisoning with medication at the index act, again in line with previous studies (Runeson *et al.*, 2016). In most cases, hospital stays were of short duration (almost 70% staying one night or less; 79.4% less than 1 week in intensive care) and 70% returned home. These observations raise several questions, notably (i) the relevance of developing specific programs targeting adolescents and young adults to prevent initial acts; (ii) the organization and feasibility of care programs offered at hospital discharge in light of a large number of individuals engaging in DSH episodes. This latter point should also take into account the recognised fact that many individuals do not attend medical appointments after DSH (Haw *et al.*, 2002), necessitating proactive interventions.

A recurrence of DSH hospitalisation occurred in 12.4% of the sample within 1 year, during the first 6 months in 75% of cases (suggesting a same unresolved crisis), and usually only once (75%). The rate of recurrence in our study was slightly lower than reported by one meta-analysis (16.3% (CI 15.1–17.7) (Carroll *et al.*, 2014)). Most recurrent acts involved a similar means as the index act, mostly drug overdose, as previously reported (Owens *et al.*, 2015).

After 1 year, 2.6% of the total sample had died. Our study confirmed that people who engage in DSH are a fragile population related to premature mortality in comparison with the general population, with an overall SMR of 7.5. SMRs even reached more than 20 in specific age groups, notably young adults.

Importantly, the main causes of death were shared between suicide, natural causes and accident as previously reported (Hawton *et al.*, 2006; Karasouli *et al.*, 2011; Bergen *et al.*, 2012b). Follow-up of patients with DSH should therefore include both somatic and psychiatric care in a holistic approach.

Suicide was the cause of death for 0.9% of the DSH sample, a result lower than the one reported by a recent meta-analysis (1.6% (CI 1.2–2.4%) (Carroll *et al.*, 2014)). The same study however noted large between-study variations in 1-year mortality rates and our findings are within the same range as several previous studies (Hawton *et al.*, 2003; Bergen *et al.*, 2012b).

Almost 63% of suicides occurred within the first 6 months following DSH, a previously highlighted critical period (Cooper *et al.*, 2005). While the risk of suicide may persist for years in some individuals who attempt suicide (Suominen *et al.*, 2004; Hawton *et al.*, 2015), follow-up studies have previously shown that this initial period is critical regarding mortality risk following DSH (Hawton *et al.*, 2015). This period may therefore represent both the most sensitive but also convenient time frame to prevent suicidal behaviour, including suicide death and DSH recurrence.

Risk factors at the time of DSH for subsequent mortality and suicide were similar as previous studies: older age, being male, use of a violent DSH means, being admitted to an intensive care unit, recurrence of DSH and being transferred to another medical department or to a psychiatric inpatient care unit. Among violent DSH, hanging/strangulation/suffocation, use of a sharp object, firearm, gases and chemical or noxious substances were all significantly associated with an increased risk of future suicide, although significant associations with other violent means (e.g. jumping from height, drowning, immolation, jumping in front of a vehicle) may not have been found due to a lack of power, frequent death before or during the index hospitalization or, again, misclassification. The supplementary suicide risk conferred by a violent *v.* non-violent methods (mainly drug overdose) has previously been reported (Bergen *et al.*, 2012a; Stenbacka and Jokinen, 2015; Olfson *et al.*, 2017; Beckman *et al.*, 2018).

In our sample, a little less than a quarter of patients were admitted in a psychiatry inpatient unit after the index act, which is higher than in some countries (e.g. 7% in the UK (Cooper *et al.*, 2013)). Admission to psychiatry likely reflects concerns by the physician about the level of short-term suicide risk or the severity of comorbid psychopathology but also available local beds and outpatient facilities. It is noteworthy that psychiatry inpatient care was not associated with a lower risk of suicide death compared with individuals who also engaged in DSH but were not hospitalised in psychiatry. These results are in line with previous studies in other countries (Hjorthøj *et al.*, 2014; Kapur *et al.*, 2015). We can only speculate on the reasons such as more severe psychopathology, inadequate care with regard to suicide risk in some psychiatric hospitals, and insufficient follow-up after hospitalisation. Discharge from psychiatric hospital is associated with a short-term risk of suicide, notably within the first 3 months (Olfson *et al.*, 2016).

Studies have shown that clinician's estimate of suicide risk is insufficient. In the present study, we confirmed several risk factors for suicide. However, if taken alone, reported risk factors are of little help at the time of index hospitalisation to identify the individuals who will die from suicide within the next year. Indeed, positive predictive values for use of violent suicidal means or admission in intensive care unit hardly reached 2%. *i.e.* less than 2% with the risk factor will die from suicide. Similar results were previously reported for the use of scales alone (Carter *et al.*,



2017), leading governmental agencies like NICE in England (NICE guidelines, 2011) to propose risk reduction to all patients following DSH.

In conclusion, this national study confirmed that people who engage in DSH are at increased risk of both short-term DSH recurrence and premature death from suicide, but also other causes. The critical period appears to be the first 6 months following DSH. Individual risk factors are of very limited utility suggesting that all patients should be followed in the aftercare of DSH as long as more specific risk factors (or combination of factors) are not identified. These results support currently developing strategies to reduce both DSH recurrence and related mortality.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S2045796019000039>

**Availability of data and materials.** Data cannot be shared to date as they include administrative non-anonymous information that can only be handed by selected institutions (here, the Ministry of Health).

**Author ORCIDs.**  F. Jollant, 0000-0001-5809-4503

**Acknowledgements.** None

**Financial support.** This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

**Conflicts of interest.** All authors have no conflict of interest to declare.

**Ethical standards.** The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

## References

- Beckman K, Mittendorfer-Rutz E, Waern M, Larsson H, Runeson B and Dahlin M (2018) Method of self-harm in adolescents and young adults and risk of subsequent suicide. *Journal of Child Psychology and Psychiatry* **59**, 948–956.
- Bergen H, Hawton K, Waters K, Ness J, Cooper J, Steeg S and Kapur N (2012a) How do methods of non-fatal self-harm relate to eventual suicide? *Journal of Affective Disorders* **136**, 526–533.
- Bergen H, Hawton K, Waters K, Ness J, Cooper J, Steeg S and Kapur N (2012b) Premature death after self-harm: a multicentre cohort study. *Lancet* **380**, 1568–1574.
- Bernal M, Haro JM, Bernert S, Brugha T, de Graaf R, Bruffaerts R, Lepine JP, de Girolamo G, Vilagut G, Gasquet I, Torres JV, Kovess V, Heider D, Neelman J, Kessler R and Alonso J (2007) Risk factors for suicidality in Europe: results from the ESEMED study. *Journal of Affective Disorders* **101**, 27–34.
- Carroll R, Metcalfe C and Gunnell D (2014) Hospital presenting self-harm and risk of fatal and non-fatal repetition: systematic review and meta-analysis. *PLoS One* **9**, e89944.
- Carter G, Milner A, McGill K, Pirkis J, Kapur N and Spittal MJ (2017) Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. *British Journal of Psychiatry* **210**, 387–395.
- Cavanagh JT, Carson AJ, Sharpe M and Lawrie SM (2003) Psychological autopsy studies of suicide: a systematic review. *Psychological Medicine* **33**, 395–405.
- Charlson ME, Pompei P, Ales KL and MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases* **40**, 373–383.
- Chung CH, Lai CH, Chu CM, Pai L, Kao S and Chien WC (2012) A nationwide, population-based, long-term follow-up study of repeated self-harm in Taiwan. *BMC Public Health* **12**, 744.
- Cooper J, Kapur N, Webb R, Lawlor M, Guthrie E, Mackway-Jones K and Appleby L (2005) Suicide after deliberate self-harm: a 4-year cohort study. *American Journal of Psychiatry* **162**, 297–303.
- Cooper J, Steeg S, Bennewith O, Lowe M, Gunnell D, House A, Hawton K and Kapur N (2013) Are hospital services for self-harm getting better? An observational study examining management, service provision and temporal trends in England. *BMJ Open* **3**, e003444.
- Fedyszyn IE, Erlangsen A, Hjorthøj C, Madsen T and Nordentoft M (2016) Repeated suicide attempts and suicide among individuals with a first emergency department contact for attempted suicide: a prospective, nationwide, danish register-based study. *Journal of Clinical Psychiatry* **77**, 832–840.
- GBD 2015 (2016) Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the global burden of disease study 2015. *Lancet* **388**, 1603–1658.
- Geulayov G, Casey D, McDonald KC, Foster P, Pritchard K, Wells C, Clements C, Kapur N, Ness J, Waters K and Hawton K (2018) Incidence of suicide, hospital-presenting non-fatal self-harm, and community-occurring non-fatal self-harm in adolescents in England (the iceberg model of self-harm): a retrospective study. *The Lancet. Psychiatry* **5**, 167–174.
- Haw C, Houston K, Townsend E and Hawton K (2002) Deliberate self harm patients with depressive disorders: treatment and outcome. *Journal of Affective Disorders* **70**, 57–65.
- Hawton K, Zahl D and Weatherall R (2003) Suicide following deliberate self-harm: long-term follow-up of patients who presented to a general hospital. *British Journal of Psychiatry* **182**, 537–542.
- Hawton K, Harriss L and Zahl D (2006) Deaths from all causes in a long-term follow-up study of 11,583 deliberate self-harm patients. *Psychological Medicine* **36**, 397–405.
- Hawton K, Bergen H, Cooper J, Turnbull P, Waters K, Ness J and Kapur N (2015) Suicide following self-harm: findings from the multicentre study of self-harm in England, 2000–2012. *Journal of Affective Disorders* **175**, 147–151.
- Hjorthøj CR, Madsen T, Agerbo E and Nordentoft M (2014) Risk of suicide according to level of psychiatric treatment: a nationwide nested case-control study. *Social Psychiatry and Psychiatric Epidemiology* **49**, 1357–1365.
- Kapur N, Steeg S, Turnbull P, Webb R, Bergen H, Hawton K, Geulayov G, Townsend E, Ness J, Waters K and Cooper J (2015) Hospital management of suicidal behaviour and subsequent mortality: a prospective cohort study. *The Lancet. Psychiatry* **2**, 809–816.
- Kapusta ND, Tran US, Rockett IR, De Leo D, Naylor CP, Niederkrotenthaler T, Voracek M, Etzersdorfer E and Sonneck G (2011) Declining autopsy rates and suicide misclassification: a cross-national analysis of 35 countries. *Archives of General Psychiatry* **68**, 1050–1057.
- Karasouli E, Owens D, Abbott RL, Hurst KM and Dennis M (2011) All-cause mortality after non-fatal self-poisoning: a cohort study. *Social Psychiatry and Psychiatric Epidemiology* **46**, 455–462.
- NICE guidelines (2011) Self-harm in over 8s: long-term management. 1–29. Available at: (<https://www.nice.org.uk/guidance/cg133>). Accessed 17 January 2019.
- Olfson M, Wall M, Wang S, Crystal S, Liu SM, Gerhard T and Blanco C (2016) Short-term suicide risk after psychiatric hospital discharge. *JAMA Psychiatry* **73**, 1119–1126.
- Olfson M, Wall M, Wang S, Crystal S, Gerhard T and Blanco C (2017) Suicide following deliberate self-harm. *American Journal of Psychiatry* **174**, 765–774.
- Owens D, Horrocks J and House A (2002) Fatal and non-fatal repetition of self-harm. *Systematic review. British Journal of Psychiatry* **181**, 193–199.
- Owens D, Kelley R, Munyombwe T, Bergen H, Hawton K, Cooper J, Ness J, Waters K, West R and Kapur N (2015) Switching methods of self-harm at repeat episodes: findings from a multicentre cohort study. *Journal of Affective Disorders* **180**, 44–51.
- Palmer BS, Bennewith O, Simkin S, Cooper J, Hawton K, Kapur N and Gunnell D (2015) Factors influencing coroners' verdicts: an analysis of verdicts given in 12 coroners' districts to researcher-defined suicides in England in 2005. *Journal of Public Health (Oxf)* **37**, 157–165.

- Payne RA, Oliver JJ, Bain M, Elders A and Bateman DN (2009) Patterns and predictors of re-admission to hospital with self-poisoning in Scotland. *Public Health* **123**, 134–137.
- Perry IJ, Corcoran P, Fitzgerald AP, Keeley HS, Reulbach U and Arensman E (2012) The incidence and repetition of hospital-treated deliberate self-harm: findings from the world's first national registry. *PLoS One* **7**, e31663.
- Runeson B, Tidemalm D, Dahlin M, Lichtenstein P and Långström N (2010) Method of attempted suicide as predictor of subsequent successful suicide: national long term cohort study. *British Medical Journal* **341**, c3222.
- Runeson B, Haglund A, Lichtenstein P and Tidemalm D (2016) Suicide risk after nonfatal self-harm: a national cohort study, 2000–2008. *Journal of Clinical Psychiatry* **77**, 240–246.
- Shepard DS, Gurewich D, Lwin AK, Reed GA and Silverman MM (2016) Suicide and suicidal attempts in the United States: costs and policy implications. *Suicide and Life Threatening Behavior* **46**, 352–362.
- Stenbacka M and Jokinen J (2015) Violent and non-violent methods of attempted and completed suicide in Swedish young men: the role of early risk factors. *BMC Psychiatry* **15**, 196.
- Suominen K, Isometsa E, Suokas J, Haukka J, Achte K and Lonnqvist J (2004) Completed suicide after a suicide attempt: a 37-year follow-up study. *American Journal of Psychiatry* **161**, 562–563.
- Tsiachristas A, McDaid D, Casey D, Brand F, Leal J, Park AL, Geulayov G and Hawton K (2017) General hospital costs in England of medical and psychiatric care for patients who self-harm: a retrospective analysis. *The Lancet. Psychiatry* **4**, 759–767.

SUPPLEMENTAL MATERIAL

Figure 1. Number of deaths per age group and gender for each of the 4 mortality causes during the one-year period following hospitalization for deliberate self-harm.

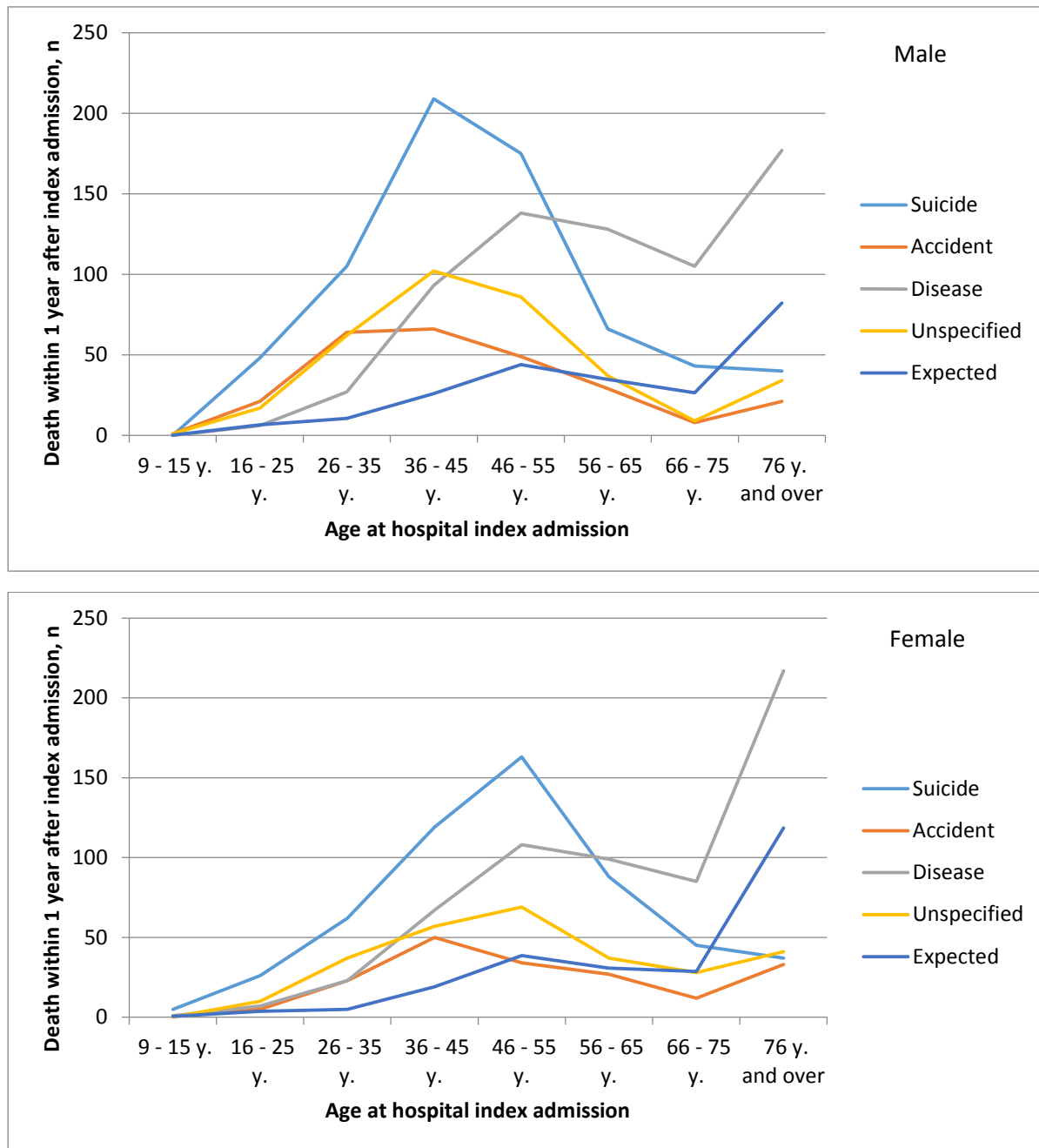
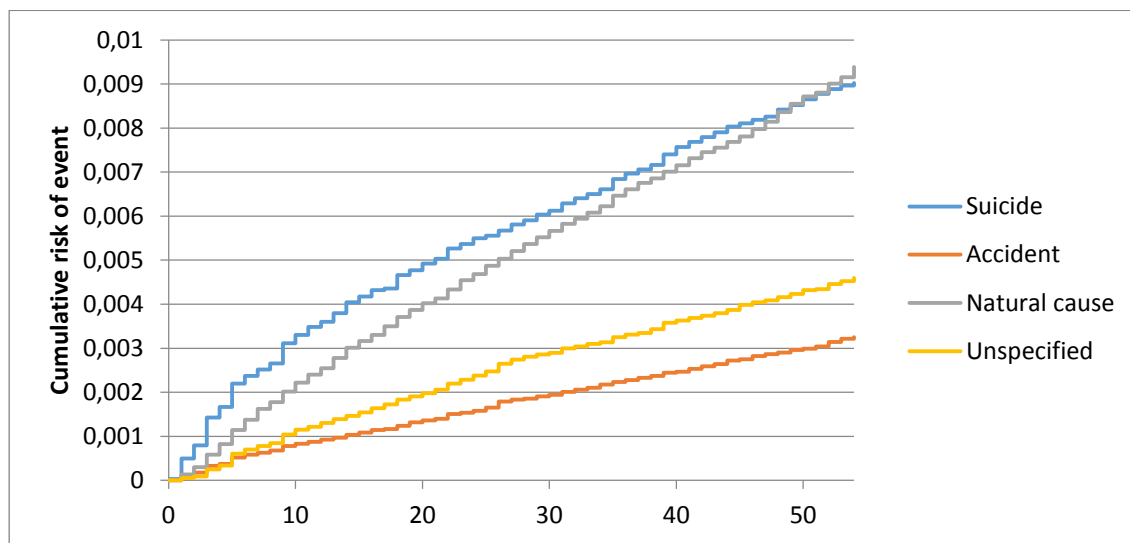




Figure 2. Cumulative risk curves for specific causes of mortality since index hospital admission for deliberate self-harm.



#### Predictive risk factors of mortality according to DSH means

Detailed examination according to DSH means showed an increased risk of overall mortality at one year in comparison to drug ingestion when the index DSH method was hanging, strangulation or suffocation (OR=1.75 95%CI [1.45-2.12]), including in relation to suicide (OR=2.05 [1.55-2.70]) and natural causes (OR=1.92 [1.41-2.62]); and cutting with sharp or blunt objects (OR=1.39 [1.19-1.62]) in relation to suicide (OR=1.35 [1.05-1.73], natural causes (OR=1.37 [1.07-1.78]) and unspecified causes (OR=1.60 [1.13-2.25]). Moreover, firearm discharge (OR=1.99 [1.23-3.20]), gases and vapors (OR=2.32 [1.30-4.13]) and ingestion of unspecified chemicals and noxious substances (OR=1.39 [1.02-1.90]) were all associated with an increased risk of future suicide, while drowning and submersion was associated with future accidental death only (OR=2.70 [1.10-6.65]). Surprisingly, use of unspecified means at index DSH tended to be associated with increased risks of death by unspecified causes one year later (OR=1.5 [1.1-2.0]). Other means (jumping from height, jumping in front of a vehicle, immolation/explosion) were not significantly associated with increased one-year mortality relative to drug ingestion.

## Discussion

Le dispositif mis en place, initialement pour l'intégration des données du programme de médicalisation du système d'information (PMSI) dans la base de données de l'Assurance maladie, de génération pour chaque résumé de séjour d'hospitalisation, d'un pseudonyme par hachage des traits d'identification des patients, apparait, dans les faits, suffisamment robuste pour réaliser le chaînage des résumés entre eux et ainsi ouvrir la possibilité de restituer de façon déidentifiée les parcours de soins hospitaliers. Le dispositif tire sa robustesse de la transversalité des référentiels de traits d'identification utilisés (numéro d'assuré social, date de naissance, code sexe), de leur accessibilité dans le contexte de soins hospitaliers, de l'utilisation d'une fonction de hachage ne générant pas sauf exception des collisions, de la normalisation des informations en entrée de la fonction de hachage par un logiciel destiné à tous les établissements de santé, enfin de la constance et de l'unicité du triplet de traits d'identification pour chaque patient, tout du moins de façon générale. La robustesse a été mesurée selon diverses méthodes, cohérence interne des informations, articulation avec la base de données de l'Assurance maladie. Cette validation est nécessaire, les informations d'identification étant collectées de façon asynchrone, sans procédure de confirmation en temps réel.

Le pseudonyme est associé à une information de positionnement chronologique relatif, propre à chaque pseudonyme, les résumés ne comportant, pour des motifs de respect de la vie privée des patients, pas de dates autres que le mois et l'année de sortie. Le positionnement chronologique relatif apporte quant à elle, une information précise au jour. La cohérence du positionnement chronologique relatif vis à vis des mois et l'année de sortie a été vérifiée. Les deux informations de chaînage, combinées avec celles des résumés PMSI, permettent, moyennant une mise en forme des données avec un logiciel de traitement des données, de restituer le parcours de soins hospitalier et de calculer différentes valeurs, délais de survenue, de réadmission en particulier, de caractériser des situations comme des épisodes de soins avec transferts entre établissements. Le logiciel diffusé aux établissements normalise la transformation des dates en délais, un guide d'instruction de recueil des informations des résumés PMSI est également transmis. Dès lors, les caractéristiques des patients, la synthèse de leur pathologie, des actes médicaux réalisés, peuvent être placés dans ce canevas de parcours hospitaliers restitués sans identification directe ni dates, et décrits avec des référentiels communs.

Les possibilités de description et d'analyse ont été illustrées par deux exemples d'études, la suite dans le mois des hospitalisations avec chirurgie, le devenir dans l'année des hospitalisations pour lésions auto-infligées. Ces études, outre la force apportée par le nombre des observations et l'étendue des établissements de santé, apportent des connaissances nouvelles en l'absence de données d'enquête spécifique : le faible effet associé au changement du mode de financement par comparaison à l'évolution induite par les modifications des techniques ; une vulnérabilité plus importante qu'admise précédemment, des patients avec lésions auto-infligées. Ces études suggèrent des développements complémentaires, par exemple de tenter de qualifier avec les informations disponibles dans les résumés PMSI le type des réadmissions après chirurgie (suite de prise en charge, complication, pathologie intercurrente, passage par les urgences). Des hypothèses physiopathologiques peuvent être interrogées comme par exemple à partir du profil des causes de décès, par âge et sexe, dans l'année qui suit une hospitalisation pour lésion auto-infligée, la similitude de la distribution des âges entre suicide et accident pouvant faire évoquer une forme de continuum ; les causes inconnues ayant le même profil, l'hypothèse de la perte du certificat de décès à l'occasion de l'enquête de police en cas de mort violente est à retenir et si ceci était avéré, augmenterait probablement les données de létalité par suicide.

Il peut être envisagé un développement ultérieur pour les deux études publiées. Outre la qualification du type de réadmission, l'analyse du moment de survenue de celle-ci peut apporter des informations utiles. En effet, il est admis que les réadmissions dans les jours qui suivent une intervention chirurgicale

sont d'avantage associées à des complications de l'intervention et que les réadmissions plus tardives voient plus fréquemment des difficultés de coordination des soins. Pour ce dernier aspect, la prise en compte des données de densité médicales ou de structures de soins autour des zones de résidence des patients peut apporter un complément d'information. Un autre angle d'analyse prendrait la perspective du développement continu de la chirurgie ambulatoire et le devenir des patients en termes de réhospitalisation. Il convient de noter que le recueil PMSI ne prévoit pas la notion de soins programmés, par exemple en fonction du délai entre prise de contact et admission effective en hospitalisation ; or l'une des caractéristiques de la chirurgie ambulatoire est d'être programmée pour bien préparer les soins post-opératoires à domicile ; une approche combinant durée de séjour, mode d'entrée, type d'intervention devra être conduite. Ce travail a été engagé et des premiers résultats présentés au congrès de l'Adelf.

L'étude de devenir des patients hospitalisés pour lésions auto-infligées peut elle aussi être étendue en la portant sur les données du système national des données de santé. Ceci permettrait de confirmer les résultats observés sur des données récentes et comportant maintenant pour l'ensemble des assurés sociaux le statut vital. Des connaissances nouvelles peuvent aussi être attendues de cette source de données : association entre profil des consultations après hospitalisation, ou de l'hospitalisation en psychiatrie avec le pronostic, estimation des risques de décès avec un recul plus important. S'agissant d'événement de fréquence peu importante, l'apport d'une base de données de ce volume devrait résulter en une contribution de forte valeur épidémiologique ; des analyses par sous-groupes de patients, notamment en fonction des âges de la vie, peuvent apporter des connaissances spécifiques, tout comme l'analyse des comorbidités et les risques compétitifs des causes de décès.

Ces exemples d'études et de leur extensions ne sont pas limitatifs, tant la diversité des pathologies prises en charge est importantes, la restitution du parcours de soins hospitalier donnant, plus qu'un descriptif ponctuel, des connaissances utiles à la pratique clinique. Le mode de collecte normalisé, exhaustif, donnent de la force. Cependant, bien que globalement robuste le dispositif n'est pas entaché de certaines faiblesses.

Pour ce qui concerne la génération du pseudonyme, des valeurs différentes peuvent correspondre au même patient lorsque celui-ci change de numéro d'assuré social, en particulier pour les enfants lors du passage à la vie adulte mais aussi suite à un usage du numéro d'assuré social des deux parents, voire de la personne qui en a la garde. Le changement de pseudonyme implique une rupture de chaînage des résumés PMSI et donc l'impossibilité de reconstitution du parcours hospitalier complet ; ceci est irrémédiable, les rapprochements entre pseudonymes non signifiant et de distribution uniforme n'étant pas possible. Cette situation survient aussi pour les patients qui perdent leur conjoint qui était leur assuré social. Il convient aussi de garder en mémoire la possibilité lors de la constitution de fichiers source dans les établissements de santé, d'erreurs d'enregistrement, d'absence d'information, de compléments ultérieurs, situations qui sont toutes à l'origine de pseudonymes différents pour un même patient. La fréquence de ces situations n'est pas uniforme, l'admission en urgence, un défaut dans le système d'information de l'établissement, dans le processus de validation des identités des patients, y étant associés. La situation inverse, le même pseudonyme valant pour deux personnes distinctes, existe structurellement dans le dispositif pour les enfants jumeaux de même sexe, ce qui est une limitation constitutive pour des études portant sur la néonatalogie en particulier.

Pour ce qui concerne l'information portée par les codes de diagnostic, deux limites peuvent être soulevées. Par nature, le codage doit être aussi synthétique que possible et ne porter que sur les pathologies effectivement prises en charge au cours de l'hospitalisation. Ces règles peuvent résulter en une description médicale insuffisamment détaillée pour répondre à des questions pourtant d'intérêt clinique, par exemple la sévérité d'une insuffisance cardiaque chronique, l'importance des séquelles

d'un accident vasculaire cérébral ; l'analyse des descriptions doit en tenir compte. La seconde limite est celle de la complétude du codage. Il est admis, pour un motif de finalité initiale du recueil PMSI de valorisation financière via un algorithme qui ne prend en compte qu'au plus quelques codes de diagnostics, qu'une part des pathologies pourtant prises en charge ne sont pas codées, l'information ne figurant pas dans la lettre de sortie ou l'effort de documentation semblant ne pas en valoir la peine ; cette situation peut avoir un impact sur les mesures obtenues, notamment pour la quantification de facteurs de risque. Des codes de diagnostics peuvent aussi figurer en excès, suite par exemple à la confusion dans les documents utilisés pour le codage, ou encore une erreur lors de l'enregistrement. Cette situation a un impact fort pour des codes de pathologies rares, donc la valeur prédictive est alors diminuée. Ceci a été objectivé lors du dénombrement des codes de pathologies à déclaration obligatoire, par comparaison aux données de veille sanitaire : de façon ponctuelle mais constante figurent des codes de rage, peste, charbon, ... . Ceci a un impact potentiel pour les études portant sur des pathologies orphelines, pourtant l'une des utilités d'un recueil exhaustif à l'échelle d'un pays.

De fait, le recueil PMSI ayant été porté par les services du ministère de la Santé à des fins de pilotage du financement des établissements de santé, voire d'outil d'encadrement des comptes publics, au moyen d'un nombre limité de groupes de forfaits, l'animation coordonnée de la précision, la complétude et l'exactitude du codage apparaissait secondaire et la défiance des cliniciens vis-à-vis de ce recueil d'autant plus grande. Cependant, la non implication des services du ministère de la santé pour une haute qualité scientifique du codage, peut aussi résulter dans une certaine neutralité dans la procédure de codage dans les établissements, qui devient une routine, sans enjeux directs pour des travaux scientifiques et les pressions que cela peut engendrer. Des pratiques de codage extrêmes peuvent cependant exister dans certains établissements, et il est de bon aloi de vérifier dans la mesure du possible, la cohérence entre établissements des résultats obtenus. Il peut sinon être regretté l'absence de suggestions de méthodes tendant à éviter les pertes d'informations au sein des établissements de santé, mais là aussi il convient de compter sur une évolution lente des pratiques et ne pas escompter qu'elles puissent changer sur ordre, les professionnels devant d'abord s'y reconnaître.

Pour installer encore d'avantage la robustesse des résultats statistiques issus de la base de données PMSI, le contenu de celle-ci peut être structurée selon les informations qu'elle comporte. La consolidation des informations autour des identifiants d'établissements de santé et dans leur évolution pluriannuelle, permet de vérifier la consistance du codage de chacun d'entre eux vis-à-vis des autorisations d'activité, et discuter de la prise en compte dans les résultats de données de séjours correspondant à des discordances identifiées mais aussi de défauts ponctuels manifestes. Une telle consolidation est en cours de constitution. Une analyse systématique des combinaisons, au sein des résumés PMSI, de codes de diagnostic et d'actes médicaux est une autre possibilité pour identifier des situations extrêmes, puis pour en discuter le maintien dans les analyses. Une approche par des techniques automatique a été présentée au congrès de l'Adelf.

Le chaînage de données médicales sans recours à des traits d'identification en clair, transformés au moyen d'algorithme de hachage en un pseudonyme non signifiant en soi dans le cadre du respect de la vie individuelle privée, peut aussi se faire à partir d'autres informations que celles utilisées pour le pseudonyme associé aux résumés PMSI. Le numéro d'inscription au répertoire national des personnes physiques et non le numéro d'assuré social pourrait être utilisé ; cependant, cette information apparaît moins accessible que ne l'est le numéro d'assuré social. L'utilisation des nom de naissances, premier prénom, date de naissance, code sexe apparaît comme ayant une valeur opérationnelle similaire à celle des clés de chaînage PMSI, un traitement de normalisation orthographique ayant toutefois été appliqué, limitant ainsi les ruptures de chaînage d'origine typographique. Ceci peut avoir toute son utilité pour des études multicentriques n'envisageant pas d'appariement ultérieur direct avec les données du système national des données de santé. Cette méthode offre en outre un circuit plus simple d'agrément, le

recours au le numéro d'assuré social étant spécifiquement encadré par la réglementation. Des collisions, même pour des noms et prénoms fréquents, apparaissent devoir être très rares. Enfin, la génération de pseudonymes peut être réalisée sur chaque trait d'identification séparément permettant ensuite d'appliquer des méthodes d'appariement probabiliste tendant à équilibrer pertes indues de chaînage et chaînage en excès.

## Conclusion

Le pseudonyme, une chaîne de caractères résultant de l'application d'une fonction de hachage au triplet d'information numéro d'assuré social, date de naissance, code sexe du patient, et l'information de position chronologique relative associés à chaque résumé d'hospitalisation produit dans le cadre du programme de médicalisation du système d'information (PMSI), apparaissent comme des informations suffisamment robustes et cohérentes pour restituer le parcours de soins hospitaliers, ceci sans avoir recours aux données d'identité des patients ni des dates des prises en charge. La robustesse est obtenue par la mise en œuvre d'éléments techniques normalisés qui prennent appui sur des informations d'utilisation commune dans les établissements de santé. Ce dispositif permet, tout en limitant les données au strict nécessaire et donc à la préservation des droits des patients à une vie individuelle privée, la réalisation d'études statistiques.

La base de données nationale PMSI, donc la finalité initiale est le pilotage du financement de l'activité de soins des établissements de santé, offre donc des possibilités de description et d'analyse des parcours de soins hospitaliers par son seul contenu (pathologies, démographie, actes médicaux, évolution) et que les caractéristiques du recueil (normalisation, systématique, exhaustif) rendent d'autant plus pertinentes. Les possibilités d'études sont très diverses, mais cependant limitées à la description synthétique des pathologies faites dans les résumés, sans détail clinique. Deux exemples d'applications sont proposés, apportant des éléments de connaissance nouveaux (évolution des pratiques, devenir des pathologies), utiles tant pour la pratique clinique que son organisation, voire la formulation d'hypothèses pouvant alimenter des travaux de recherche ultérieurs. Le pseudonyme, conçu initialement comme connecteur des résumés PMSI aux données de l'Assurance maladie, apparaît comme remplissant effectivement ce rôle, ouvrant alors sur des possibilités de description et d'analyse des parcours étendues aux actes de soins non exclusivement hospitaliers. L'accès à la base de données est encadré réglementairement.

## Références bibliographiques

- Bouillaguet C. Modes opératoires pour les fonctions de hachage. In Principes et algorithmes cryptographiques, Université de Lille, 2016 53-8
- Brams L. Carte à mémoire dans le secteur de la santé – Rapport de synthèse des travaux du comité scientifique de la mission carte communication santé. Ministère de la santé, 1992. 30pp
- Carmille R. La mécanographie dans les administrations. 2<sup>nde</sup> édition Recueil Sirey, 1941
- CESSI, centre d'étude des sécurités du système d'information. Traitement des données personnelles de santé à des fins d'évaluation ou d'analyse des activités de soins – Fonction d'occultation des informations nominatives. Caisse nationale de l'Assurance maladie, 2002. 14pp.
- Chaperon J, Trehony A. Aperçu historique des méthodes employées pour la mise en œuvre de l'enquête statistique de morbidité des hôpitaux généraux publics. Santé, sécurité sociale : statistiques et commentaires, ministère des affaires sociales et de la solidarité nationale, 1980 n°4 : 29-37.
- CNAMTS, caisse nationale d'assurance maladie des travailleurs salariés. Classification commune des actes médicaux. Journal officiel de la République française, 11 mars 2005.
- CNIL, commission nationale de l'informatique et des libertés. XXème rapport d'activité pour l'année 1999. La Documentation française, Paris, 2000. 360pp.
- Cour, Cours des comptes. Actions du ministère des affaires sociales en matière d'informatique hospitalière. In Rapport au président de la République, 1992, Paris.
- Dusserre L, Quantin C, Bouzelat H. A one way public key cryptosystem for the linkage of nominal files in epidemiological studies. MEDINFO 95, RA Greenes, HE Peterson, DJ Protti (editors), Elsevier Science Publishers (North-Holland), 1995, p 644-647.
- Fuchs J. Etat des lieux et perspectives sur l'utilisation des bases de données médico-administratives de l'Assurance maladie en matière de pharmaco-épidémiologie. Thèse de pharmacie, Université Claude Bernard Lyon I, 2015. 106pp.
- Grmek MD, Rey R. Physiologie et médecines des Lumières. Dix-Huitième Siècle. 1991 ; 23(1) : 5-8
- Gutmann P, Naccache D, Palmer CC. When hashes collide [applied cryptography]. IEEE Security & Privacy, 2005; 3(3): 68-71
- Karnofsky DA, Burchenal JH. In: Evaluation of chemotherapeutic agents. MacLeod CM, editor. New York: Columbia University Press; 1949. The clinical evaluation of chemotherapeutic agents in cancer; pp. 191–205.
- Laferriere N, Tenailon A, Saltiel JC, Smagghe A, Chicou FJ, Chretien J, Portos JL. Le questionnaire médical – Principe, construction et évaluation, application. INSERM, 1978, Paris. 81pp.
- Lawson EH, Ko CY, Louie R, Han L, Rapp M, Zingmond DS. Linkage of a clinical surgical registry with medicare inpatient claims data using indirect identifiers. Surgery. 2013;153(3) : 423-30.
- Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA 1993 ; 270 : 2957-2963.
- Linder FE, Krueger DE. Uniform hospital abstract : minimum basis data set. US department of health, education and welfare 1972, 17p.
- Ministère de l'emploi et de la solidarité, direction de l'hospitalisation et de l'organisation des soins. Circulaire DHOS-PMSI n° 106 relative au chaînage des séjours en établissements de santé. Bulletin officiel, 2001(13), texte n°893.
- Ministère de l'emploi et de la solidarité, direction de l'hospitalisation et de l'organisation des soins. Circulaire DHOS/E 3 n° 2001-625 du 19 décembre 2001 relative à la mise en œuvre du PMSI-psychiatrie, à titre expérimental. Bulletin officiel, 2002a(7), texte n°547

- Ministère de l'emploi et des solidarités, direction de la sécurité sociale. Protocole inter-régimes relatif au système national d'information inter-régimes de l'assurance maladie. Bulletin officiel, 2002b(14), texte n°1343
- Ministère de la santé. Arrêté portant création d'un traitement automatisé d'informations dans le cadre d'une étude de faisabilité de la mise en place d'un système d'information sur la mortalité post-hospitalière. Journal officiel de la République française, 2004, 5 mai : 8010
- Montastruc JL, Lugardon S, Desbœuf K, Fernet P, Lapeyre-Mestre M. Utilisation de la méthode capture-recapture pour l'évaluation de la fréquence des effets indésirables « graves » médicamenteux : l'expérience du CHU de Toulouse Bull. Acad. Natle Méd. 2008 ; 192 (2) : 421-431
- Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic Linkage of Vital Records. Science 1959 ; 130 : 954-9
- OMS, organisation mondiale de la santé. Classification statistique internationale des maladies et des problèmes de santé connexes – 10<sup>ème</sup> révision. Genève, 1993. Tomes 1, 2 et 3.
- OMS, organisation mondiale de la santé. Classification CIM-10 des troubles mentaux et des troubles du comportement – descriptions cliniques et directives pour le diagnostic ; critères diagnostiques pour la recherche. Genève, 1994. 248pp.
- Ordre national des médecins. Le médecin du département d'information médicale dans les établissements de santé – Réglementation et déontologie. Masson, Paris, 2000. 101p.
- Preneel B. The first 30 years of cryptographic hash function and the NIST SHA-3 competition. In Pieprzyk J (eds) Topics in cryptology – CT-RSA 2010. Lecture notes in computer science, Springer, Berlin : 1-14.
- Quantin C, Bouzelat H, Allaert FA, Benhamiche AM., Faivre J, Dusserre L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. Methods of Information in Medicine. 1998a; 37:271-7.
- Quantin C, Bouzelat H, Allaert FA, Benhamiche AM., Faivre J, Dusserre L. How to ensure data security of an epidemiological follow-up : quality assessment of an anonymous record linkage procedure, International Journal of Medical Informatics. 1998b; 49:117-22.
- Quantin C, Gouyon B, Allaert FA, Dusserre L, Cohen O. Méthodologie pour le chaînage de données sensibles tout en respectant l'anonymat : application au suivi des informations médicales. Courrier des statistiques, 2005 ;113-114 :15-26 ; Journal de la SFdS 2005 ;146(3) :15-26.
- Quantin C, Benzenine E, Allaert FA, Guesdon M, Gouyon Jb, Riandey B. Epidemiological and Statistical Secured Matching in France. Statistical Journal of the IAOS 30 (2014a) 255–61
- Quantin C, Cottenet J, Vuagnat A, Prunet C, Mouquet MC, Fresson J, Blondel B. Qualité des données périnatales issues du PMSI : comparaison avec l'état civil et l'enquête nationale périnatale 2010 ; Journal de Gynécologie Obstétrique et Biologie de la Reproduction 2014b ; 43 : 680-90
- Robelet M. Les médecins placés sous observation. Mobilisation autour du développement de l'évaluation médicale en France. Politix, 1999 ; 12(2) : 71-97.
- Roger FH. The minimum basic data set for hospital statistics in the EEC – Review of availability and comparability. Commission of the European Communities, 1981 Bruxelles 151p.
- Scheuren F. Linking Health Records: Human Rights Concerns. in National Research Council. 1999. Record Linkage Techniques-1997: Proceedings of an International Workshop and Exposition. Washington, DC: The National Academies Press.p 404-26
- Trouillet JL, Scheimberg A, Vuagnat A, Fagon JY, Chastre J, Gibert C. Long-term outcome and quality of life of patients requiring multidisciplinary intensive care unit admission after cardiac operations. J. Thorac. Cardiovasc. Surg. 1996; 112: 926-34.



- U.S. Department of commerce - National Institute of Standards and Technology. Secure Hash Standard. Federal information processing standards publication, 1995; 180-1; 21pp.
- Vallet O, Villeminot S, Gasquet I, Falissard B. Élaboration d'un outil de mesure de la mortalité post hospitalière. Ministère de la santé, de la famille et des personnes handicapées - Direction de la recherche, des études, de l'évaluation et des statistiques. 2004 Document de travail n°40, tomes 1 et 2. 117pp.
- Valois J, Abrial JR, Cohen S. Description d'un système de gestion en temps réel du dossier médical. Grenoble Médico-Chirurgical 1970; 8 : 301-14.
- Weill A, Païta P, Tuppin P, Fagot JP, Neumann A, Simon D, Ricordeau P, Montastruc JL, Allemand H. Benfluorex and valvular heart disease: a cohort study of a million people with diabetes mellitus. *Pharmacoepidemiology and drug safety*. 2010; 19(12) : 1256-62
- Zins M. Bases de données administratives et médico-administratives – Appariement dans un contexte épidémiologique. Présentation à École d'Eté Méditerranéenne d'Information en Santé, Corte Juillet 2013

## Table des matières

Introduction	3
Repères relatifs à la constitution des données médicales hospitalières	6
Données du programme de médicalisation du système d'information, PMSI	6
Chaînage des résumés PMSI	9
Travaux préparatoires : analyse de la cohérence des données PMSI	15
Comparaison de résultats issus des données du PMSI à ceux d'autres sources	15
Description des pseudonymes dits « clés de chaînage » PMSI	18
Valeur opérationnelle des clés de chaînage PMSI comparée à celle de clés obtenues avec d'autres informations d'identification	20
Apport de la base Amphi, précurseur du Système national des données de santé	21
Vérification formelle de la mise en œuvre de l'algorithme SHA-1	24
Description des parcours de soins hospitaliers avec le PMSI : matériel et méthodes	26
Résultats : deux exemples d'application	30
Évolution sur la période 2002-2012 du taux de réadmission après une hospitalisation pour chirurgie	30
Devenir à un an des patients admis pour tentative de suicide	41
Discussion	54
Conclusion	58
Références bibliographiques	59
Table des matières	62
Annexes	63
Facsimilé de la carte perforée utilisée pour le recensement des activités professionnelles en 1941	63
Diagramme de l'algorithme SHA-1, exemple de programme assembleur	64
Publications et communications	70
Nombre cumulé de consultations depuis les sites internet des revues des articles	73



## Diagramme de l'algorithme SHA-1, exemple de programme assembleur

Restitution à partir de U.S. Department of commerce - National Institute of Standards and Technology. Secure Hash Standard. Federal information processing standards publication, 1995

L'information à hacher est découpée en blocs de 512 bits et placée en mémoire dans 16 mots de 4 octets  
Pour le dernier bloc, à la fin de l'information en entrée est ajouté le bit 1 puis des 0 pour compléter (padding)  
les 16 mots de 4 octets, la longueur de l'information en entrée en bits est indiquée à la toute fin du dernier bloc

↓  
A partir des ces 16 mots de 4 octets est calculée une séquence de 64 mots de 4 octets en mémoire de la façon suivante:  
Permutation binaire circulaire de 1 à gauche de (mot situé à-3 ou\_exclusif mot situé à-8 ou\_exclusif mot situé à-14 ou\_exclusif mot situé à-16)  
Est ainsi définie une séquence de 80 mots de 4 octets qui vont être incorporés l'un après l'autre dans la fonction suivante

↓  
Initialisation en mémoire de 5 mots de 4 octets: A=0x67452301, B=0xEFCDAB89, C=0x98BADCFE, D=0x10325476, E=0xC3D2E1F0

↓  
calcul d'un 6<sup>ème</sup> mot qui incorpore à chaque étape le mot suivant de la séquence de 80 mots  
T= permutation binaire circulaire de 5 à gauche de A + f(B,C,D) + E + mot à incorporer + constante  
La fonction et la constante changent à chaque sous séquence de 20 mots incorporés  
fonction: (B et C) ou\_inclusif (complément de B et D), B ou\_exclusif C ou\_exclusif D  
(B et C) ou\_inclusif (B et D) ou\_inclusif (C et D), B ou\_exclusif C ou\_exclusif D  
constante: 0x5A827999, 0x6ED9EBA1, 0x8F1BBCDC, 0xCA62C1D6

↓  
Une permutation entre les 5 mots est ensuite réalisée:  
E=D; D=C; C= permutation binaire circulaire de 30 à gauche de B; B=A; A=T

↓  
A l'issue de la séquence d'incorporation des 80 mots de la séquence d'entrée, les valeurs d'initialisation sont ajoutées respectivement à chaque mot A, B, C, D, E

↓  
S'il n'y a plus de bloc à traiter, la signature (résultat produit pas SHA-1) correspond à la concaténation des 5 mots de 4 octets.  
Sinon un nouveau bloc de 512 bits est placé en mémoire et incorporé (ce n'est pas nécessaire pour les clé de chaînage PMSI)

Exemple de programme assembleur : écriture directement en langage machine, les codes binaires sont présentés par des abréviations qui présentent le sens des opérations élémentaires réalisables avec les microprocesseurs

```
/*
 * SHA-1 hash in x86 assembly
 *
 * Copyright (c) 2014 Project Nayuki
 * https://www.nayuki.io/page/fast-sha1-hash-implementation-in-x86-assembly
 *
 * (MIT License)
 * Permission is hereby granted, free of charge, to any person obtaining a copy of
 * this software and associated documentation files (the "Software"), to deal in
 * the Software without restriction, including without limitation the rights to
 * use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
 * the Software, and to permit persons to whom the Software is furnished to do so,
 * subject to the following conditions:
 * - The above copyright notice and this permission notice shall be included in
 *   all copies or substantial portions of the Software.
 * - The Software is provided "as is", without warranty of any kind, express or
 *   implied, including but not limited to the warranties of merchantability,
 *   fitness for a particular purpose and noninfringement. In no event shall the
 *   authors or copyright holders be liable for any claim, damages or other
 *   liability, whether in an action of contract, tort or otherwise, arising from,
 *   out of or in connection with the Software or the use or other dealings in the
 *   Software.
 */
```

```

/* void sha1_compress(uint32_t state[5], const uint8_t block[64]) */
.globl sha1_compress
sha1_compress:
/*
 * Storage usage:
 *   Bytes  Location  Description
 *   4     eax       SHA-1 state variable A
 *   4     ebx       SHA-1 state variable B
 *   4     ecx       SHA-1 state variable C
 *   4     edx       SHA-1 state variable D
 *   4     ebp       SHA-1 state variable E
 *   4     esi       Temporary for calculation per round
 *   4     edi       (During key schedule rounds) base address of
block array argument (read-only); (during hash rounds) temporary for
calculation per round
 *   4     esp       x86 stack pointer
 *   320   [esp+ 0]   Key schedule of 80 x 4 bytes
 *   4     [esp+320] Caller's value of ebp
 *   4     [esp+324] Caller's value of edi
 *   4     [esp+328] Caller's value of esi
 *   4     [esp+332] Caller's value of ebx
 */

#define LOADSCHEDULE(i) \
    movl    (i*4)(%edi), %esi; \
    bswapl  %esi; \
    movl    %esi, (i*4)(%esp);

#define SCHEDULE(i) \
    movl    ((i-3)*4)(%esp), %esi; \
    xorl    ((i-8)*4)(%esp), %esi; \
    xorl    ((i-14)*4)(%esp), %esi; \
    xorl    ((i-16)*4)(%esp), %esi; \
    roll    $1, %esi; \
    movl    %esi, (i*4)(%esp);

#define ROUND0(a, b, c, d, e, i) \
    movl    %c, %esi; \
    xorl    %d, %esi; \
    andl    %b, %esi; \
    xorl    %d, %esi; \
    ROUNDTAIL(a, b, e, i, 0x5A827999)

#define ROUND1(a, b, c, d, e, i) \
    movl    %b, %esi; \
    xorl    %c, %esi; \
    xorl    %d, %esi; \
    ROUNDTAIL(a, b, e, i, 0x6ED9EBA1)

#define ROUND2(a, b, c, d, e, i) \
    movl    %c, %esi; \
    orl     %d, %esi; \
    andl    %b, %esi; \
    movl    %c, %edi; \
    andl    %d, %edi; \
    orl     %edi, %esi; \
    ROUNDTAIL(a, b, e, i, 0x8F1BBCDC)

#define ROUND3(a, b, c, d, e, i) \
    movl    %b, %esi; \
    xorl    %c, %esi; \
    xorl    %d, %esi; \

```

```

        ROUNDTAIL(a, b, e, i, 0xCA62C1D6)

#define ROUNDTAIL(a, b, e, i, k) \
        addl  %esi, %e;          \
        movl  %a, %edi;          \
        roll  $5, %edi;          \
        addl  %edi, %e;          \
        addl  (i*4)(%esp), %e;   \
        addl  $k, %e;            \
        roll  $30, %b;

/* Save registers */
pushl  %ebx
pushl  %esi
pushl  %edi
pushl  %ebp

/* Load arguments */
movl   20(%esp), %esi /* state */
movl   24(%esp), %edi /* block */
movl   0(%esi),  %eax /* a */
movl   4(%esi),  %ebx /* b */
movl   8(%esi),  %ecx /* c */
movl  12(%esi),  %edx /* d */
movl  16(%esi),  %ebp /* e */

/* 80 words of key schedule */
subl   $320, %esp
LOADSCHEDULE( 0)
LOADSCHEDULE( 1)
LOADSCHEDULE( 2)
LOADSCHEDULE( 3)
LOADSCHEDULE( 4)
LOADSCHEDULE( 5)
LOADSCHEDULE( 6)
LOADSCHEDULE( 7)
LOADSCHEDULE( 8)
LOADSCHEDULE( 9)
LOADSCHEDULE(10)
LOADSCHEDULE(11)
LOADSCHEDULE(12)
LOADSCHEDULE(13)
LOADSCHEDULE(14)
LOADSCHEDULE(15)
SCHEDULE(16)
SCHEDULE(17)
SCHEDULE(18)
SCHEDULE(19)
SCHEDULE(20)
SCHEDULE(21)
SCHEDULE(22)
SCHEDULE(23)
SCHEDULE(24)
SCHEDULE(25)
SCHEDULE(26)
SCHEDULE(27)
SCHEDULE(28)
SCHEDULE(29)
SCHEDULE(30)
SCHEDULE(31)
SCHEDULE(32)
SCHEDULE(33)

```

SCHEDULE (34)  
SCHEDULE (35)  
SCHEDULE (36)  
SCHEDULE (37)  
SCHEDULE (38)  
SCHEDULE (39)  
SCHEDULE (40)  
SCHEDULE (41)  
SCHEDULE (42)  
SCHEDULE (43)  
SCHEDULE (44)  
SCHEDULE (45)  
SCHEDULE (46)  
SCHEDULE (47)  
SCHEDULE (48)  
SCHEDULE (49)  
SCHEDULE (50)  
SCHEDULE (51)  
SCHEDULE (52)  
SCHEDULE (53)  
SCHEDULE (54)  
SCHEDULE (55)  
SCHEDULE (56)  
SCHEDULE (57)  
SCHEDULE (58)  
SCHEDULE (59)  
SCHEDULE (60)  
SCHEDULE (61)  
SCHEDULE (62)  
SCHEDULE (63)  
SCHEDULE (64)  
SCHEDULE (65)  
SCHEDULE (66)  
SCHEDULE (67)  
SCHEDULE (68)  
SCHEDULE (69)  
SCHEDULE (70)  
SCHEDULE (71)  
SCHEDULE (72)  
SCHEDULE (73)  
SCHEDULE (74)  
SCHEDULE (75)  
SCHEDULE (76)  
SCHEDULE (77)  
SCHEDULE (78)  
SCHEDULE (79)

```
/* 80 rounds of hashing */  
ROUND0 (eax, ebx, ecx, edx, ebp, 0)  
ROUND0 (ebp, eax, ebx, ecx, edx, 1)  
ROUND0 (edx, ebp, eax, ebx, ecx, 2)  
ROUND0 (ecx, edx, ebp, eax, ebx, 3)  
ROUND0 (ebx, ecx, edx, ebp, eax, 4)  
ROUND0 (eax, ebx, ecx, edx, ebp, 5)  
ROUND0 (ebp, eax, ebx, ecx, edx, 6)  
ROUND0 (edx, ebp, eax, ebx, ecx, 7)  
ROUND0 (ecx, edx, ebp, eax, ebx, 8)  
ROUND0 (ebx, ecx, edx, ebp, eax, 9)  
ROUND0 (eax, ebx, ecx, edx, ebp, 10)  
ROUND0 (ebp, eax, ebx, ecx, edx, 11)  
ROUND0 (edx, ebp, eax, ebx, ecx, 12)  
ROUND0 (ecx, edx, ebp, eax, ebx, 13)
```

ROUND0 (ebx, ecx, edx, ebp, eax, 14)  
ROUND0 (eax, ebx, ecx, edx, ebp, 15)  
ROUND0 (ebp, eax, ebx, ecx, edx, 16)  
ROUND0 (edx, ebp, eax, ebx, ecx, 17)  
ROUND0 (ecx, edx, ebp, eax, ebx, 18)  
ROUND0 (ebx, ecx, edx, ebp, eax, 19)  
ROUND1 (eax, ebx, ecx, edx, ebp, 20)  
ROUND1 (ebp, eax, ebx, ecx, edx, 21)  
ROUND1 (edx, ebp, eax, ebx, ecx, 22)  
ROUND1 (ecx, edx, ebp, eax, ebx, 23)  
ROUND1 (ebx, ecx, edx, ebp, eax, 24)  
ROUND1 (eax, ebx, ecx, edx, ebp, 25)  
ROUND1 (ebp, eax, ebx, ecx, edx, 26)  
ROUND1 (edx, ebp, eax, ebx, ecx, 27)  
ROUND1 (ecx, edx, ebp, eax, ebx, 28)  
ROUND1 (ebx, ecx, edx, ebp, eax, 29)  
ROUND1 (eax, ebx, ecx, edx, ebp, 30)  
ROUND1 (ebp, eax, ebx, ecx, edx, 31)  
ROUND1 (edx, ebp, eax, ebx, ecx, 32)  
ROUND1 (ecx, edx, ebp, eax, ebx, 33)  
ROUND1 (ebx, ecx, edx, ebp, eax, 34)  
ROUND1 (eax, ebx, ecx, edx, ebp, 35)  
ROUND1 (ebp, eax, ebx, ecx, edx, 36)  
ROUND1 (edx, ebp, eax, ebx, ecx, 37)  
ROUND1 (ecx, edx, ebp, eax, ebx, 38)  
ROUND1 (ebx, ecx, edx, ebp, eax, 39)  
ROUND2 (eax, ebx, ecx, edx, ebp, 40)  
ROUND2 (ebp, eax, ebx, ecx, edx, 41)  
ROUND2 (edx, ebp, eax, ebx, ecx, 42)  
ROUND2 (ecx, edx, ebp, eax, ebx, 43)  
ROUND2 (ebx, ecx, edx, ebp, eax, 44)  
ROUND2 (eax, ebx, ecx, edx, ebp, 45)  
ROUND2 (ebp, eax, ebx, ecx, edx, 46)  
ROUND2 (edx, ebp, eax, ebx, ecx, 47)  
ROUND2 (ecx, edx, ebp, eax, ebx, 48)  
ROUND2 (ebx, ecx, edx, ebp, eax, 49)  
ROUND2 (eax, ebx, ecx, edx, ebp, 50)  
ROUND2 (ebp, eax, ebx, ecx, edx, 51)  
ROUND2 (edx, ebp, eax, ebx, ecx, 52)  
ROUND2 (ecx, edx, ebp, eax, ebx, 53)  
ROUND2 (ebx, ecx, edx, ebp, eax, 54)  
ROUND2 (eax, ebx, ecx, edx, ebp, 55)  
ROUND2 (ebp, eax, ebx, ecx, edx, 56)  
ROUND2 (edx, ebp, eax, ebx, ecx, 57)  
ROUND2 (ecx, edx, ebp, eax, ebx, 58)  
ROUND2 (ebx, ecx, edx, ebp, eax, 59)  
ROUND3 (eax, ebx, ecx, edx, ebp, 60)  
ROUND3 (ebp, eax, ebx, ecx, edx, 61)  
ROUND3 (edx, ebp, eax, ebx, ecx, 62)  
ROUND3 (ecx, edx, ebp, eax, ebx, 63)  
ROUND3 (ebx, ecx, edx, ebp, eax, 64)  
ROUND3 (eax, ebx, ecx, edx, ebp, 65)  
ROUND3 (ebp, eax, ebx, ecx, edx, 66)  
ROUND3 (edx, ebp, eax, ebx, ecx, 67)  
ROUND3 (ecx, edx, ebp, eax, ebx, 68)  
ROUND3 (ebx, ecx, edx, ebp, eax, 69)  
ROUND3 (eax, ebx, ecx, edx, ebp, 70)  
ROUND3 (ebp, eax, ebx, ecx, edx, 71)  
ROUND3 (edx, ebp, eax, ebx, ecx, 72)  
ROUND3 (ecx, edx, ebp, eax, ebx, 73)  
ROUND3 (ebx, ecx, edx, ebp, eax, 74)  
ROUND3 (eax, ebx, ecx, edx, ebp, 75)



```
ROUND3(ebp, eax, ebx, ecx, edx, 76)
ROUND3(edx, ebp, eax, ebx, ecx, 77)
ROUND3(ecx, edx, ebp, eax, ebx, 78)
ROUND3(ebx, ecx, edx, ebp, eax, 79)

addl    $320, %esp

/* Save updated state */
movl    20(%esp), %esi
addl    %eax, 0(%esi)
addl    %ebx, 4(%esi)
addl    %ecx, 8(%esi)
addl    %edx, 12(%esi)
addl    %ebp, 16(%esi)

/* Restore registers */
popl    %ebp
popl    %edi
popl    %esi
popl    %ebx
retl
```

## Publications et communications

### Publications dans des revues disposant d'un comité de lecture

- Quantin C, Cottenet J, Vuagnat A, Prunet C, Mouquet MC, Fresson J, Blondel B. Qualité des données périnatales issues du PMSI : comparaison avec l'état civil et l'enquête nationale périnatale 2010. *J Gynecol Obstet Biol Reprod* 2013 ; 43 : 680-90.
- Pierron A, Revert M, Goueslard K, Vuagnat A, Cottenet J, Benzenine E, Fresson J, Quantin C. Évaluation de la qualité métrologique des données du programme de médicalisation du système d'information (PMSI) en périnatalité : étude pilote réalisée dans 3 CHU. *Rev Epidemiol Sant Pub* 2015 ; 63 :237-46.
- Chaussade H, Uçkay I, Vuagnat A, Druon J, Gras G, Rosset P, Lipsky BA, Bernard L. Antibiotic therapy duration for prosthetic joint infections treated by Debridement and Implant Retention (DAIR): Similar long-term remission for 6 weeks as compared to 12 weeks. *Int J Infect Dis.* 2017; 63 : 37-42.
- Vuagnat A, Yilmaz E, Roussot A, Rodwin V, Gadreau M, Bernard A, Creuzot-Garcher C, Quantin C. Did case-based payment influence surgical readmission rates in France? A retrospective study. *BMJ Open* 2018 Feb 1; 8(2)
- Boisguérin B, Toutlemonde F, Vuagnat A. Characteristics of patient management in emergency departments. *Soins.* 2018; 63 : 16-19
- Naouri D, El Khoury C, Vincent-Cassy C, Vuagnat A, Schmidt J, Yordanov Y; French Society of Emergency Medicine Evaluation and Quality Committee. The French Emergency National Survey: A description of emergency departments and patients in France. *PLoS One* 2018 Jun 14;13(6).
- Vuagnat A, Jollant F, Abbar M, Hawton K, Quantin C. Recurrence and mortality 1 year after hospital admission for non-fatal self-harm: a nationwide population-based study. *Epidemiol Psychiatr Sci.* 2019 Feb 18:1-10.
- Naouri D, Ranchon G, Vuagnat A, Schmidt J, El Khoury C, Yordanov Y; French Society of Emergency Medicine. Factors associated with inappropriate use of emergency departments: findings from a cross-sectional national study in France. *BMJ Qual Saf.* 2020; 29: 449-64.

### Publications dans des ouvrages ou des revues sans comité de lecture

- Vuagnat A. Les parcours de soins hospitaliers. p 132-3. In *Le panorama des établissements de santé - Edition 2012*, Boisguérin B et Minodier C *edrs*. Direction de la recherche, des études, de l'évaluation et de statistiques, Paris, 2012. 172 pp.
- Vuagnat A. Les urgences hospitalières, qu'en sait-on ? p 11-28. In *Le panorama des établissements de santé - Edition 2013*, Boisguérin B et Brilhault G *edrs*. Direction de la recherche, des études, de l'évaluation et de statistiques, Paris, 2014. 162 pp.
- Yilmaz E, Vuagnat A. Tarification à l'activité et réadmission. *Économie et statistique*, Insee, 2015, n°475-476 : 71-87.
- Vuagnat A. Effets période et génération mesurables dans les statistiques de décès par suicide. p 417-23. In 2<sup>ème</sup> rapport de l'Observatoire national du suicide Fourcade N et von Lenep F *edrs*, Paris, 2016. 480 pp.
- Vuagnat A. Relation entre hospitalisation en MCO pour tentative de suicide et suicide. p 424-32. In 2<sup>ème</sup> rapport de l'Observatoire national du suicide Fourcade N et von Lenep F *edrs*, Paris, 2016. 480 pp.
- Vuagnat A, Toutlemonde F. Histoire des urgences et des évaluations de l'adéquation à leur recours. *Actualité et dossier en santé publique* n° 109, décembre 2019 : 6-9
- Vuagnat A. Les conduites suicidaires en France : toujours un problème de santé publique. *Rev Prat.* 2020;70: 32-7

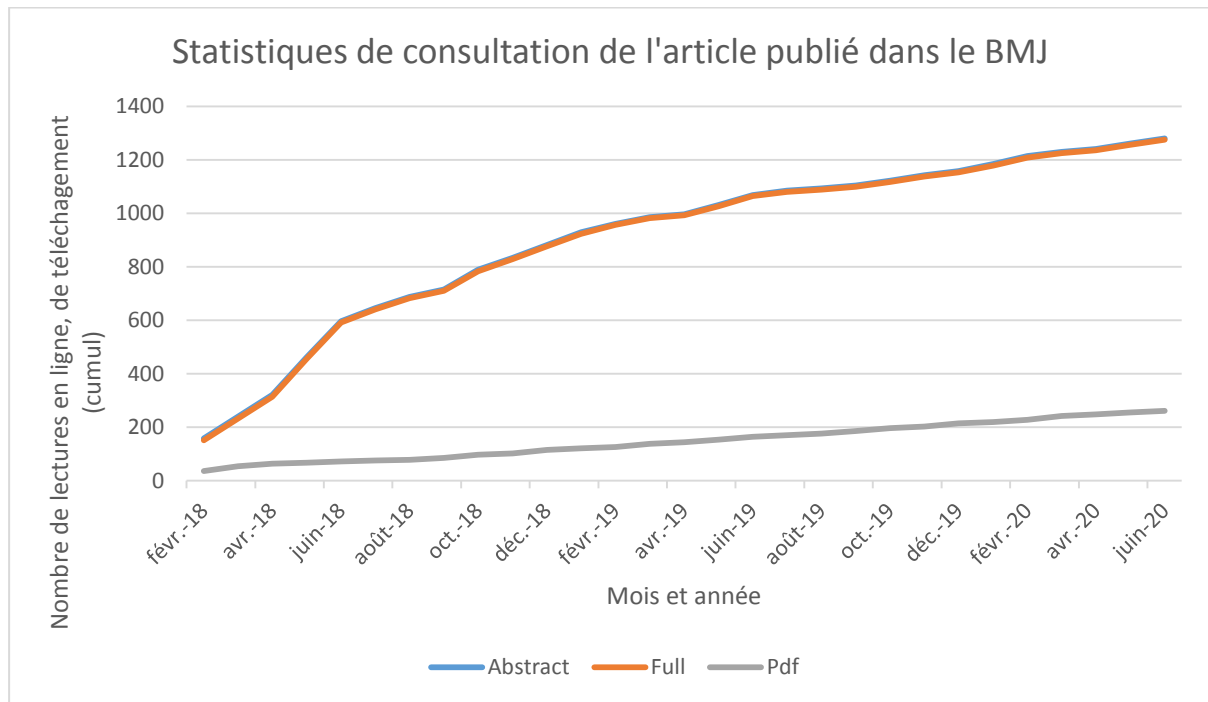
## Communications aux congrès scientifiques

- Vuagnat A, Bensadon M. Génération des clés de chaînage des séjours hospitaliers : comparaison des caractéristiques opératoires du hachage de deux jeux de données différents. Travail présenté au congrès de l'Association des Épidémiologistes De Langue Française, Saint Malo, 2008.
- Bernard L, Vuagnat A, Assal M, Hoffmeyer P, Uçkay I. Diagnostic microbiologique de 154 ostéites : utilité de 2 prélèvements profonds successifs de fistule. Travail présenté aux journées nationales d'infectiologie, Lyon, 2009.
- Vuagnat A, Bensadon M. Qualité du codage du score IGS2 à l'admission en réanimation de patients adultes, un impact de la mise en œuvre de la tarification à l'activité ? - Base PMSI nationale 2005-2009. Travail présenté au congrès de la Société de Réanimation de Langue Française, Paris, 2011.
- Valdelièvre H, Coquelet F, Evain F, Gisquet E, Morin L, Oberlin P, Vuagnat A. Prise en charge des patients nécessitant des soins palliatifs dans les établissements de santé. Travail présenté aux congrès de l'Association des Épidémiologistes De Langue Française, Dijon, 2012.
- Roussot A, Vuagnat A, Benzenine E, Amat-Roze JM, Combier E, Quantin C. Accessibilité de la maternité la plus proche en Bourgogne, comparaison de deux méthodes de calcul d'itinéraires. Travail présenté aux journées Évaluation, Management, Organisation, Information en Santé, Nancy, 2013.
- Quantin, C, Cottenet J, Vuagnat A, Benzenine E, Fresson J, Prunet C, Moquet MC, Grémy I, Gouyon JB, Blondel B. Validité des données périnatales issues du PMSI : comparaison avec les données de l'enquête nationale périnatale 2010. Travail présenté aux journées Évaluation, Management, Organisation, Information en Santé, Nancy, 2013.
- Vuagnat A, Mas R. Détermination du nombre de mesures d'hospitalisation sans consentement au moyen du RIM-P. Travail présenté aux journées Évaluation, Management, Organisation, Information en Santé, Nancy, 2013.
- Minodier C, Boisguérin B, Valdelièvre H, Vuagnat A. Enquête un jour donné sur les services de urgences. Travail présenté au congrès de la Société Française de Médecine d'Urgence, Paris, 2013.
- Vuagnat A. Statut vital à un an des patients hospitalisés en réanimation – une utilisation de la base AMPHI. Travail présenté au congrès de l'Association des Épidémiologistes De Langue Française, Paris, 2014.
- Yilmaz E, Vuagnat A. Évolution entre 2005 et 2012 du taux de réadmission à 30 jours après un séjour chirurgical : bases PMSI nationales. Travail présenté au congrès de l'Association des Epidémiologistes De Langue Française, Paris, 2014.
- Pierron A, Revert M, Goueslard K, Vuagnat A, Cottenet J, Benzenine, Fresson J, Quantin C. Évaluation de la qualité métrologique des données du Programme de médicalisation du système d'information (PMSI) en périnatalité. Étude pilote réalisée dans les CHU de Dijon, Port-Royal et Nancy. Travail présenté aux journées Évaluation, Management, Organisation, Information en Santé, Nancy, 2015.
- Vuagnat A. Quelles traces dans le PMSI-MCO du défaut de natalité observé pendant les guerres mondiales ? Travail présenté aux journées Évaluation, Management, Organisation, Information en Santé, Nancy, 2015.
- Ricroch L, Vuagnat A. Temps, parcours et délais aux urgences : résultats de l'enquête nationale. Travail présenté aux journées Évaluation, Management, Organisation, Information en Santé, Nancy, 2015.
- De Peretti C. L, Vuagnat A. Décès à un an après une hospitalisation pour accident vasculaire cérébral en 2008 ou 2009. Travail présenté au congrès de la Société Française Neuro-Vasculaire, Paris, 2015.
- De Peretti C. L, Vuagnat A. Patients hospitalisés pour infarctus cérébral en 2008 ou 2009 : admission en soins intensifs et mortalité à un an. Travail présenté au congrès de l'Association des Épidémiologistes De Langue Française, Dijon, 2016.

- Lefebvre-Hoang I. L, Vuagnat A. Typage de l'ensemble des unités médicales de court-séjour : quelle information sur le fonctionnement des établissements de santé ? Travail présenté au congrès de l'Association des Épidémiologistes De Langue Française, Dijon, 2016.
- Vuagnat A. Tarification à l'activité : quel impact sur les réadmissions à l'hôpital ? Travail présenté à la Journée du Journal de Gestion et d'Économie Médicale, Paris, 2017.
- Yilmaz E, Vuagnat A. Développement de la chirurgie ambulatoire : impact sur les ré-hospitalisations. Travail présenté au congrès de l'Association des Épidémiologistes De Langue Française, Montpellier, 2018.
- Vuagnat A., Jollant F., Abbar M., Hawton K., Quantin C. Mortalité à un an post-hospitalisation pour lésion auto-infligée: une étude nationale française. Travail présenté au congrès de l'Association des Épidémiologistes De Langue Française, Montpellier, 2018.
- Doutréline M., Leduc A., Nguyen D.-P., Vuagnat A. Snds2vec, représentations continues pour les concepts médicaux du Système national des données de santé. Travail présenté au congrès de l'Association des Épidémiologistes De Langue Française, Paris, 2020.

**Nombre cumulé de consultations depuis les sites internet des revues des articles de la section résultats, à la date du 28 juin 2020 (consultable via l'onglet metrics du site internet de chaque revue)**

Vuagnat A, Yilmaz E, Roussot A, Rodwin V, Gadreau M, Bernard A, Creuzot-Garcher C, Quantin C. Did case-based payment influence surgical readmission rates in France? A retrospective study. *BMJ Open*

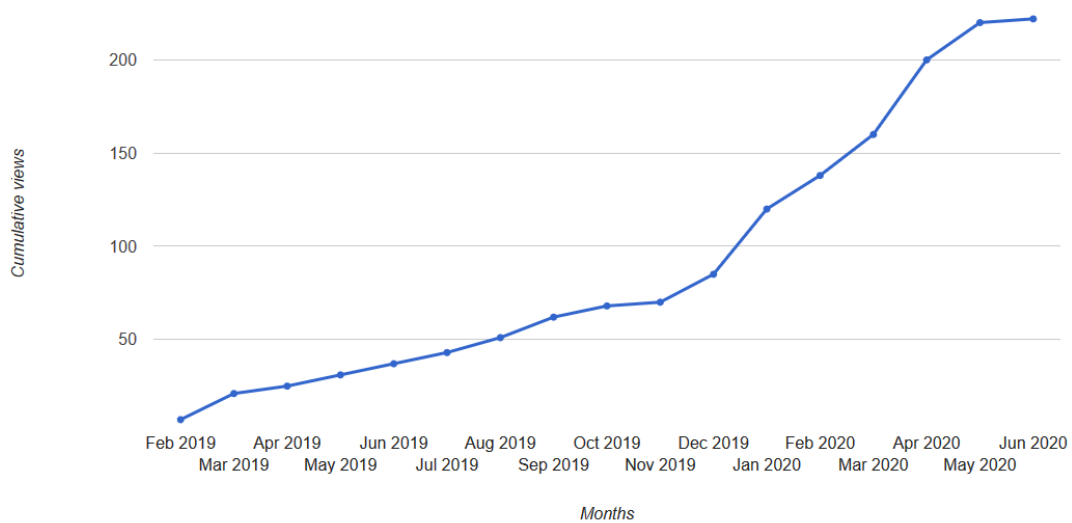


Vuagnat A, Jollant F, Abbar M, Hawton K, Quantin C. Recurrence and mortality 1 year after hospital admission for non-fatal self-harm: a nationwide population-based study. *Epidemiol Psychiatr Sci.*

Full text views

Total number of HTML

Full text views reflects the number of PDF downloads, PDFs sent to Google Drive, Dropbox and Kindle and HTML full text views.



**Titre :** Peut-on utiliser les données du programme de médicalisation du système d'information pour décrire les parcours de soins ?

**Mots-clés :** PMSI, chaînage, parcours de soins

**Résumé :** Le programme de médicalisation du système d'information (PMSI) organise dans les établissements de santé un recueil systématique et standardisé pour toute hospitalisation : les pathologies et leur prise en charge sont codées sous la forme d'un résumé informatique auquel sont associées des informations de chaînage : une chaîne de caractère non signifiante résultant du hachage des informations numéro d'assuré social, date de naissance, code sexe du patient, et une position chronologique relative sont produites. Sous réserve de la constance et de l'uniformité de la collecte des informations source comme de l'algorithme de hachage, les informations de chaînage associées aux résumés PMSI d'un même patient permettent ainsi, à priori, de reconstituer, de façon anonyme, des parcours de soins hospitaliers. L'objectif du travail est de vérifier la réalité de cet à priori dans les données de la base nationale du PMSI. Une

première étape a permis de conforter la robustesse des informations. Cette robustesse est obtenue par la mise en œuvre d'éléments techniques qui prennent appui sur des informations d'utilisation commune dans les établissements de santé. Moyennant une mise en forme des données, il est possible de restituer sans recours à l'identité des patients ni aux dates de prises en charge, informations ne figurant pas dans la base de données nationale du PMSI, des parcours de soins hospitaliers cohérents. Deux exemples d'études décrivant et analysant le devenir des patients hospitalisés (suite dans le mois qui suit un séjour pour chirurgie, devenir dans l'année des hospitalisations pour lésions auto-infligées) illustrent l'apport de connaissances que peut apporter cette source d'information. L'accès à la base de données est encadré par la réglementation.

**Title :** An evaluation of linked administrative data in the perspective of hospital outcome measurement

**Key words :** claim data, data linkage, hospital outcome

**Abstract :** The French medical information system program (PMSI) defines the item and format all hospitals and clinics in France have to use for claim data: diagnoses, medical procedures, patient, structure and stay characteristics, this for any hospital stay. These medical records abstracts are linked by means of a salted SHA-1 hash of social security number, date of birth and gender. Hash is generated by a unique software, used by all hospitals and clinics, that also calculates a delay in days since, for privacy reasons, no dates of hospital admission are mentioned in the abstracts. The data are gathered in a national database. This collection of linked abstracts designed for funding, could also be used to describe medical history and hospital outcome. The aim of the present work is to assess

the quality of the linkage information and its potential contribution to medical knowledge. The first step was to assess quality of linkage information. This was found appropriate and consistent as a result of uniform generating process that relied on well know and standardized information. The second step was to shape the database in order to derive hospital outcome. Finally, this was used in to settings, readmission after surgery, one-year follow-up after non-fatal self-harm. To conclude this national linked claim database appears as suited for hospital outcome measurement. Access to the database has to comply with regulation defined by the committee on informatics and civil rights (CNIL).