

QUALITÉ DES DONNÉES DE L'ENQUÊTE NATIONALE SUR LES CONGÉS D'HÔPITAL

Thomas McLemore et Robert Pokras¹

RÉSUMÉ

Depuis 1965, le National Center for Health Statistics réalise l'enquête nationale sur les congés d'hôpital (* National Hospital Discharge Survey +ou NHDS), qui est une enquête nationale avec échantillon probabiliste sur les congés accordés aux patients des hôpitaux non fédéraux de courte durée ou généraux. Un grand volet du projet de remaniement de la NHDS en 1988 a consisté à exploiter des données électroniques tirées des systèmes d'information des organismes de services et des États. Le présent document donnera un aperçu de l'évolution de cette enquête et de ce remaniement de 1988. Il examinera les méthodes d'enquête à la lumière des questions de collecte et de traitement des données d'extraction *manuelle + et des données *automatisées +. Les auteurs examineront les méthodes d'évaluation de la qualité et de la précision générales des données de la NHDS pour l'un et l'autre des modes de collecte de ces données. Parmi ces méthodes, on doit ranger celles qui garantissent que les données reçues respectent les normes établies et que l'information extraite est traitée et codée suivant des règles strictes de contrôle de qualité. Ils présenteront enfin ces dernières dans le contexte des questions et des constatations des études plus générales consacrées à la qualité des ensembles de données administratives hospitalières.

MOTS CLÉS : Données sur les patients hospitalisés; Contrôle de la qualité

1. INTRODUCTION

Le National Center for Health Statistics (NCHS) est chargé d'évaluer l'état de santé de la population américaine. C'est là que se retrouvent les statistiques de l'état civil des États-Unis. Son portefeuille de collecte de données comprend une enquête nationale sur la nutrition et les examens médicaux et une enquête nationale par interviews des ménages. Le NCHS compte aussi une famille d'enquêtes nationales par échantillon probabiliste portant collectivement le nom * Enquête nationale sur les services de santé (* National Health Care Survey +ou NHCS) où on recueille des données auprès des établissements de soins et autres et des fournisseurs de services sur le territoire américain (McLemore, 2000). L'objectif premier de la NHCS est de mesurer l'utilisation du système de prestation de services de santé aux États-Unis. Ses volets sont les suivants :

Enquête nationale sur les soins médicaux ambulatoires (NAMCS) - visites de cabinets de médecin;
Enquête nationale sur les soins médicaux ambulatoires en milieu hospitalier (NHAMCS) - visites de services des urgences et de cliniques externes en milieu hospitalier;
Enquête nationale sur les soins de soins infirmiers (NNHS) - soins infirmiers;
Enquête nationale sur les soins dans les foyers et les centres d'accueil (NHHCS) - services de santé dans ces établissements;
Enquête nationale sur les services chirurgicaux ambulatoires (NSAS) - visites de services chirurgicaux ambulatoires hospitaliers ou autonomes;
Enquête nationale sur les congés d'hôpital (NHDS) - soins hospitaliers en clinique interne.

On peut se renseigner sur le NCHS et la NHCS au site Web de cet organisme à l'adresse <http://www.cdc.gov/nchs/>.

Toutes ces enquêtes sont nationales, et chacune a de nombreuses particularités, d'où des différences de conception statistique, d'exécution d'enquête et de contrôle de qualité. Le présent document donnera un aperçu

¹Division of Health Care Statistics, National Center for Health Statistics, 6525, chemin Belcrest, pièce 952, Hyattsville, Maryland 20782.

du contrôle de la qualité sous ses divers aspects dans le cas de l'enquête nationale sur les congés d'hôpital, et ce, dans le contexte de la conception et de la réalisation de cette enquête.

En 1965, le NCHS lançait l'enquête nationale sur les congés d'hôpital en vue de recueillir et de diffuser des renseignements sur l'utilisation en clinique interne des hôpitaux non fédéraux de courte durée aux États-Unis. L'importance d'une mesure des soins aux patients hospitalisés a fait de la NHDS la première enquête qu'ait menée le NCHS sur la prestation de services médicaux. Les soins en hospitalisation demeurent un grand volet du système de prestation de services de santé. La NHDS a eu lieu tous les ans depuis le début. Au nombre des variables recueillies par cette enquête, on compte les caractéristiques des patients (âge, sexe, race, région et code ZIP), des hôpitaux (taille, propriété et lieu) et des événements (dates d'entrée et de sortie, type et état d'admission, source prévue de paiement, état de sortie d'hôpital et mesures posthospitalisation), ainsi que 1 à 7 diagnostics et 0 à 4 actes médicaux selon l'International Classification of Diseases, 9th revision, Clinical Modifications (ICD-9-CM).

Les données de la NHDS sont mises à la disposition du public. Les dernières données disponibles se rapportent à 1999. Les données de 2000 pourront être consultées vers la fin de 2001. Les organismes fédéraux, les hôpitaux, les universités, la presse, les fabricants de produits médicaux, les États et les particuliers font usage des données de la NHDS, qui se retrouvent dans une diversité d'applications, qu'il s'agisse de politique de la santé, de recherche en services de santé, d'établissement d'objectifs nationaux de santé, d'épidémiologie, de recherche en santé publique ou de modélisation de marché. Une bonne source de renseignements sur la NHDS est son site Web <http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm>, où on pourra se reporter à une bibliographie des études spécialisées qui exploitent les données de cette enquête.

2. CONCEPTION DE L'ENQUÊTE

Dans les 35 ans d'évolution de la NHDS, on a appliqué deux plans d'enquête. De 1965 à 1987, on a fait un échantillonnage à deux degrés. Il y avait d'abord un échantillon probabiliste national d'hôpitaux et ensuite un sous-échantillon de congés (patients hospitalisés) des établissements participants (Simmons, 1970). C'est la période où l'enquête reposait presque entièrement sur une extraction manuelle des données des dossiers médicaux.

On a repensé la NHDS en 1988 et prévu un échantillonnage à trois degrés portant sur les unités primaires, à savoir les régions statistiques métropolitaines ou les groupes de comtés, les hôpitaux de ces secteurs et les congés des hôpitaux participants (Dennison, 2000). D'un point de vue opérationnel, un grand volet de l'initiative était l'inclusion de données puisées aux sources électroniques disponibles dans les États, les entreprises commerciales, les associations hospitalières et les divers hôpitaux.

On a effectué une analyse de comparaison des estimations des deux plans d'échantillonnage (Haupt, 1992). Les auteurs de cette analyse ont utilisé les données recueillies les trois premiers mois de 1988 par l'un et l'autre de ces plans d'enquête. Ils n'ont relevé aucune différence statistique pour la plupart des estimations des variables de l'enquête, mais ils ont aussi constaté des différences pour quelques catégories de diagnostics (cataractes, syndrome d'alcoolodépendance, etc.) et selon le nombre de lits des établissements hospitaliers.

La population visée par la NHDS actuelle est celle des congés des hôpitaux non * institutionnalisés +sans les hôpitaux fédéraux, militaires et des anciens combattants dans les 50 États et le district de Columbia. Les hôpitaux de courte durée (où les patients séjournent en moyenne moins de 30 jours) et les établissements à vocation générale (médicale ou chirurgicale) pour l'ensemble de la population ou les enfants en particulier (sans égard à la durée d'hospitalisation) sont pris en compte. Un autre critère d'admissibilité est que l'hôpital compte six lits et plus avec préposés en clinique interne.

En 1988, la base de sondage de la NHDS était formée des hôpitaux compris dans la base de données du SMG sur le marché hospitalier et qui, en août 1987, avaient commencé à recevoir des patients en clinique interne. L'échantillon d'hôpitaux a été actualisé en 1991, 1994 et 1997 en fonction des hôpitaux ayant ouvert leurs portes ou changé leur admissibilité depuis la dernière mise à jour. Chaque année, l'échantillon comprend quelque 500 établissements hospitaliers, dont 470 environ participent à l'enquête et fournissent des données annuelles sur un nombre approximatif de 300 000 congés à des hôpitaux de l'échantillon. Il y a environ 300 hôpitaux qui produisent manuellement des données sur 75 000 congés et 170 par voie électronique sur 225 000 congés en gros.

On applique des procédures de contrôle de qualité des données produites manuellement tout au long de l'enquête, et notamment aux étapes de l'échantillonnage de dossiers, de l'extraction des données, de la réception et du contrôle des documents, du codage médical et de la saisie de l'information. Les données produites électroniquement passent par une suite de mesures automatiques et manuelles de contrôle de qualité. Les dossiers communiqués tant manuellement qu'électroniquement font l'objet de vérifications informatiques rigoureuses. Il y a imputation des valeurs manquantes de sexe et d'âge, après quoi on procède à une pondération des dossiers, c'est-à-dire qu'on calcule des valeurs de pondération qui sont ensuite attribuées aux divers dossiers à des fins d'estimation. La dernière mesure de contrôle de qualité consiste en une comparaison systématique des estimations établies et de celles des années antérieures. Les mesures en question sont décrites plus loin.

3. COLLECTE MANUELLE DE DONNÉES

3.1 Aperçu

Le Census Bureau a été l'agent de collecte manuelle des données de la NHDS depuis sa création en 1965. Il comprend un bureau central et 12 bureaux régionaux. En collaboration avec le NCHS, le personnel de son bureau central a conçu la procédure d'exploitation d'enquête, rédigé, imprimé et distribué tous les manuels et formules utilisés sur le terrain aux fins de cette enquête, modifié ces documents en fonction des variables nouvellement introduites ou modifiées et surveillé en général les opérations de collecte manuelle. Le personnel des bureaux régionaux est chargé des opérations quotidiennes d'enquête, de la formation des nouveaux agents de terrain et de la fourniture des formules et autres documents de l'enquête aux préposés des hôpitaux. Les grandes activités sur le terrain sont l'initiation à l'enquête des hôpitaux échantillonnés, l'échantillonnage des congés et l'extraction des données des dossiers médicaux.

La collecte manuelle de données peut être primaire, parallèle ou par imprimés d'ordinateur. Dans le premier cas, les préposés aux archives médicales des hôpitaux échantillonnent les congés aux fins de l'enquête et tirent les données recherchées des dossiers médicaux sélectionnés pour les reporter sur formule d'enquête. C'est le mode privilégié, puisque ces préposés connaissent mieux leurs dossiers et leurs données médicaux que le personnel du Census Bureau. Dans ce cas, ils reçoivent des guides d'opérations de terrain et des consignes détaillées. Ils sont formés par les agents du Census Bureau aux tâches d'échantillonnage et d'extraction de données. En 1999, 30 % environ de tous les hôpitaux à production manuelle dans le cadre de la NHDS ont appliqué cette méthode primaire de collecte.

On recourt à la méthode parallèle lorsque des hôpitaux sont peu désireux ou incapables de faire accomplir les tâches d'échantillonnage et d'extraction par leur propre personnel. Un représentant sur le terrain (RT) du Census Bureau échantillonne alors les congés et fait en sorte que les préposés aux archives médicales mettent les dossiers médicaux à sa disposition à des fins d'extraction. Il tire lui-même les données des dossiers. On met tout en œuvre pour choisir des RT expérimentés qui connaissent bien la terminologie médicale et le travail avec les archivistes médicaux. Environ la moitié des hôpitaux à production manuelle ont appliqué cette méthode de collecte pour la NHDS de 1999.

Il y a enfin collecte * manuelle + par fourniture d'imprimés d'ordinateur pour les congés échantillonnés. On parle alors de collecte manuelle, car les données de ces imprimés doivent ensuite être entrées à la main. Les hôpitaux qui appliquent cette méthode présentent leur définition des variables codées (1 = homme, 2 = femme, etc.) pour que l'entrée des données puisse se faire correctement. L'informatisation des données des patients hospitalisés a fait des imprimés d'ordinateur quelque chose de plus courant dans la NHDS. Cette méthode de collecte a figuré pour environ 8 % dans la collecte manuelle de 1988 et pour plus de 20 % dans celle de 1999. La principale mesure de contrôle de qualité pour les imprimés est que ceux-ci font l'objet d'un examen d'acceptabilité générale par le personnel du NCHS. On se trouve à écarter les imprimés dont on ne peut extraire les données sans ambiguïté.

3.2 Échantillonnage de congés pour la collecte manuelle de données

On sélectionne des congés d'hôpital par échantillonnage aléatoire systématique où les intervalles sont fondés sur le plan statistique d'enquête dressé par le NCHS. On se reporte habituellement à des listes quotidiennes ou mensuelles de congés. On sélectionne les congés ayant certains chiffres terminaux de matricule de dossier médical. On utilise parfois un numéro d'admission ou de facturation ou un autre numéro de patient. S'il n'y a pas de matricules de patient qui se prêtent à l'échantillonnage, on sélectionne les congés en prenant un numéro de départ au hasard pour ensuite parcourir numériquement une liste de congés, puis sélectionner tout k^e congé. Dans certains établissements hospitaliers, on échantillonne par les derniers chiffres des matricules de dossier médical dans les fichiers informatiques internes.

Les données nécessaires d'identification des congés de l'échantillon sont reportées sur une liste d'échantillon. On dresse une telle liste pour l'échantillon de dossiers de chaque mois. On y trouve des indications comme les matricules de dossier médical et les dates de congé, ce qui permet de relever les dossiers médicaux à mettre en extraction de données. On garde continus les plans d'échantillonnage exigeant un passage numérique de liste de congés (comme nous l'avons décrit plus haut) en repérant sur la liste d'échantillon le point terminal de l'échantillonnage de la période précédente.

4. CONTRÔLE DE LA QUALITÉ DES DONNÉES MANUELLES

4.1 Opérations sur le terrain

Il y a deux grandes opérations d'intérêt sur le terrain, à savoir l'échantillonnage des congés et l'extraction de données des dossiers médicaux sélectionnés. Au NCHS, on contrôle la qualité de l'échantillonnage en comparant dans chaque hôpital le nombre observé de dossiers échantillonnés au nombre prévu tous les mois et en examinant le chiffre terminal des matricules des dossiers médicaux sélectionnés. Le mode d'échantillonnage est conçu de sorte que le nombre de congés mensuellement échantillonnés par l'hôpital se situe entre des valeurs limites de 5 et 50. Si le nombre de congés sélectionnés se trouve hors de cet intervalle, on marque le tout pour complément d'examen. Il y a communication des indications de mauvais échantillonnage au Census Bureau en vue d'une correction au niveau des établissements hospitaliers.

On veille principalement à une extraction sûre des données des dossiers médicaux en demandant autant que possible à des préposés qualifiés aux archives médicales dans les établissements d'extraire les renseignements des dossiers échantillonnés et de les reporter sur formule d'enquête NHDS, ce qui laisse à ceux qui connaissent les dossiers de l'hôpital et les données qu'ils renferment le soin d'effectuer la collecte primaire. S'il est impossible de recourir à la méthode primaire, les représentants de terrain du Census Bureau procèdent eux-mêmes à cette extraction. Aux fins de la NHDS, on les choisit selon leur expérience de ce type de renseignements et leur donne une formation spécialisée qui garantira la qualité de leur travail.

Longtemps, on a prévu dans la NHDS une procédure de réextraction de données d'un échantillon de dossiers médicaux à des fins de surveillance des travaux du personnel de terrain. On ne le fait plus, parce que cette réextraction est impossible dans les hôpitaux à production automatisée et que les résultats ne pourraient être obtenus en temps utile. On a souvent constaté que, lorsqu'on disposait enfin des résultats, il y avait déjà de nouveaux préposés dans les hôpitaux ou de nouveaux RT du Census Bureau qui faisaient l'extraction manuelle.

Des établissements hospitaliers, les données de collecte manuelle parviennent à un des 12 bureaux régionaux du Census Bureau, où des agents contrôlent la circulation et l'exhaustivité des relevés de données. Les listes d'échantillon et les relevés sont ensuite envoyés au sous-traitant chargé du traitement des données, lequel enregistre l'information et en suit le cheminement par une opération de réception-contrôle aux fins de cette vérification de circulation et d'exhaustivité. Il s'agit en outre de voir s'il n'y a pas de dossiers en double ou de congés manquants. De ces opérations sont issus des lots d'environ un millier de relevés, ce qui aide à surveiller tout le processus de codage médical, de saisie des données et de contrôle de la qualité.

4.2 Codage médical et saisie des données

Les consignes de terrain aux fins de la NHDS disent que les données sur les diagnostics et les actes médicaux doivent être reportées sous forme narrative sur formule d'enquête. Du personnel qualifié code les diagnostics et les actes de ces formules en se reportant à l'ICD-9-CM. On se trouve à coder un diagnostic au moins et sept au plus pour chaque relevé de l'échantillon. Il y a également codage de quatre actes de chirurgie ou de diagnostic au maximum. À l'entrée des codes dans l'ordinateur, on les vérifie à l'aide d'une liste de codes valides fournie chaque année par le NCHS. Mise à jour tous les ans, cette liste reçoit des ajouts et perd ses codes périmés. On peut ainsi corriger les entrées non valides par un réexamen de la formule d'enquête.

On surveille le codage médical et la saisie des données au moyen d'un échantillon indépendant à 10 % de chaque lot de 1 000 dossiers. Un second préposé au codage code indépendamment les relevés de l'échantillon à 10 % et un responsable du codage règle les divergences. Ajoutons que, dans le cas des nouveaux codeurs, il y a vérification à 100 % de leurs trois premiers lots et à 20 % des lots qui suivent pendant leur première année d'activité. Les lots où on relève des taux d'erreur de 5 % pour les données médicales ou de 0,5 % pour les données non médicales (saisie des données) sont rejetés et intégralement remis en codage. Le taux général d'erreur des dossiers de l'an 2000 en codage manuel a respectivement été de 0,7 % pour les diagnostics définitifs, de 0,5 % pour les actes de chirurgie et de diagnostic et de 0,2 % pour les données non médicales. Au terme de l'entrée des données et du contrôle de qualité à ce stade, on envoie l'ensemble de dossiers au NCHS à des fins de vérification, d'estimation et de traitement complémentaire.

5. COLLECTE AUTOMATISÉE DE DONNÉES

Un aspect important de la refonte de 1988 de la NHDS a été l'intégration de méthodes de collecte des données sur les patients hospitalisés qui existent sous une forme électronique. Si on a activement privilégié la production électronique et s'est employé à recueillir les données sous cette forme, ce n'est pas par seul changement de conception de l'enquête, mais aussi par de grandes modifications de l'exploitation quotidienne de la NHDS avec des contrôles de qualité à exercer tout au long des opérations.

L'idée d'une collecte de données sur la prestation de services de santé à l'aide des systèmes électroniques déjà en place a été sanctionnée au NCHS dans les années 1970 dans le cadre de ce qu'on a appelé le système coopératif de statistique de la santé (* Cooperative Health Statistics System +ou CHSS). Ce régime relevait de la loi américaine de 1974 sur la recherche en services de santé, la statistique de la santé et les bibliothèques médicales (* Health Services Research, Health Statistics, and Medical Libraries Act +, Public Law 93-353). On y mettait nettement l'accent sur une collecte nationale de données par une collaboration * de bas en haut +. Dans ce régime, les données seraient recueillies une fois, traitées initialement au niveau des États et communiquées au palier fédéral sous une forme lisible par machine en vue de la production de données nationales (Task Force,

1977). C'est ainsi que, dans les années 1970, le NCHS a financé des projets pilotes de collecte de données sur les congés d'hôpital dans 12 États. Le Département de la Santé, de l'Éducation et du Bien-être social a mis fin aux activités CHSS du NCHS en février 1979 (Final Report, 1980).

Dans une évaluation générale (Lissler, 1977) de la NHDS au milieu de la décennie 1970, on devait recommander une triple base de sondage pour cette enquête, c'est-à-dire (1) les hôpitaux relevant du CHSS, (2) les hôpitaux non visés par le CHSS qui faisaient partie d'un système privé d'extraction de données et (3) tous les autres hôpitaux pour lesquels aucune source particulière de données n'était disponible. Pour pouvoir appliquer un plan d'enquête qui prévoyait notamment une collecte automatisée, on a dû aborder les deux grandes questions de la disponibilité et de la qualité des données de ces sources. Les auteurs de l'évaluation sont parvenus à la conclusion que tous les éléments d'information de la NHDS B sauf celui de l'état matrimonial B étaient amplement disponibles à des sources automatisées.

Des études menées vers la fin des années 1970 sur la fiabilité des données d'extraction sur les congés d'hôpital (Institute of Medicine, 1977) et des données NHDS (Institute of Medicine, 1980) indiquent que les données sur les congés recueillies auprès des hôpitaux par services d'extraction respectaient et excédaient même parfois les normes de qualité qu'avait toujours cultivées le NCHS. Plus récemment encore, on a relevé des signes d'amélioration de la qualité générale des fichiers de données sur les congés, ceux-ci étant devenus importants dans la détermination des remboursements du régime d'assurance-maladie et d'autres régimes de prestations (Fisher, 1992; Green, 1993; Hsia, 1992).

Le NCHS a financé deux études consacrées aux aspects opérationnels de la collecte électronique de données auprès des services commerciaux d'extraction de données. La première de ces études (Duggar, 1981) lancée en 1980 a porté sur l'opportunité d'acheter et d'exploiter l'information des services d'extraction de données sur les congés aux fins de la NHDS. Une des constatations faites est que l'achat d'information était susceptible de largement diminuer sans l'éliminer le besoin d'une collecte directe. En 1981, environ 60 % des hôpitaux américains étaient abonnés à des services d'extraction de données sur les congés dont la clientèle était formée de 10 hôpitaux et plus. Les systèmes en question ne couvraient pas l'univers des hôpitaux, d'où l'impossibilité d'en tirer un échantillon représentatif du réseau hospitalier américain.

Dans le cadre d'une seconde étude (Miller, 1984), le NCHS a mis en route en 1984 un projet témoin visant à l'élaboration de méthodes d'acquisition, de traitement et de validation de qualité des données sur les congés qui existaient sous une forme électronique. En application de ce nouveau contrat, on a recueilli des données automatisées auprès de services commerciaux d'extraction auxquels étaient abonnés 76 hôpitaux participant à la NHDS. Les opérations d'estimation qui ont porté sur les données manuelles et automatisées de la NHDS ont donné des résultats comparables. L'étude a également démontré que les services d'extraction de données pouvaient s'en tenir aux échéances nécessaires de livraison des données de l'enquête. Fort des résultats de cette seconde étude, le NCHS a commencé à exploiter les données automatisées dans la NHDS de 1985.

De 1985 à 1987, le NCHS a fait l'acquisition de données automatisées existantes sur les patients hospitalisés dans un certain nombre d'établissements visés, et ce, en vertu d'un contrat passé avec le Center for Health Policy Studies (CHPS), qui a alors procuré au NCHS un fichier-recensement de tous les congés de 78, 77 et 74 hôpitaux selon les années considérées. Le NCHS a échantillonné les dossiers de ces fichiers suivant le plan d'échantillonnage adopté pour la collecte manuelle de données.

Depuis la refonte de 1988, le recours à une collecte automatisée s'est accru. On obtient aujourd'hui des données automatisées de quelque 170 hôpitaux tous les ans. De plus, les règles de conception prévoyaient un plus grand nombre de congés échantillonnés dans les hôpitaux qui communiquaient leurs données sous une forme lisible par machine.

6. CONTRÔLE DE LA QUALITÉ DES DONNÉES AUTOMATISÉES

Il y a deux aspects importants à tout ce qui est contrôle de la qualité de données automatisées. Il y a d'abord un contrôle de l'exhaustivité et de la qualité des données de fichiers électroniques et ensuite des vérifications informatiques. Nous décrivons plus loin les vérifications machine destinées aux fichiers manuels et automatisés. Les modes de contrôle de fichiers électroniques ont évolué. Nous dépeindrons à grands traits les méthodes employées dans le cadre de la NHDS.

Il s'agit d'une enquête à participation volontaire. Le NCHS n'est pas habilité à obliger les établissements hospitaliers à participer à la NHDS, ni à imposer la transmission de données électroniques aux fins de cette enquête en fixant la structure ou la forme. Nous recommandons une définition de fichiers avec des procédures de traitement et des structures de variables pour une telle transmission. Il reste que, soucieux de voir le taux de réponse des hôpitaux demeurer le plus haut possible, le NCHS accepte les fichiers électroniques sur divers supports (cartouches et bobines de bande, cédéroms, disquettes et fichiers joints de courrier électronique), en ASCII ou en EBCDIC, selon un éventail de dispositions et de facteurs de groupage, pour des périodes variables (trimestre, semestre et année) et, souvent, avec des structures uniques de variables. Qui plus est, il accepte les fichiers de données automatisées couvrant la population ou un échantillon de congés. Les fichiers-recensement font état de tous les congés d'un hôpital, et le NCHS se charge alors de l'échantillonnage. Les fichiers-échantillon présentent des congés préalablement échantillonnés par la source automatisée selon les spécifications du NCHS.

Tous les ans, ce dernier se voit communiquer environ 2 millions de congés sous une forme électronique par quelque 35 sources. Il reçoit ainsi une soixantaine de fichiers. Chacun de ces fichiers passe par diverses étapes de traitement. D'abord, il doit être matériellement exploitable au sens électronique du terme, sinon on doit demander un fichier de remplacement. S'il ne respecte pas notre structure de présentation, on le reformate selon une disposition type et une structure uniforme de codage de l'ensemble des variables. On l'examine à ce stade pour repérer et retrancher les dossiers en double pour un hôpital. On vérifie indépendamment le nombre de dossiers que devait envoyer chaque établissement hospitalier en demandant à chacun quel a été son nombre mensuel de congés et de sorties après accouchement.

On vérifie tous les dossiers pour y relever les variables erronées ou manquantes, les inexactitudes et les valeurs aberrantes. On soumet en outre la plupart des variables de chaque hôpital à une analyse détaillée. Pour les codes de diagnostics et d'actes, la procédure de contrôle produit une liste de codes non valides ICD-9-CM, ainsi qu'une répartition de l'ensemble des congés selon les chapitres descriptifs de cette nomenclature. Pour ces variables et toutes les autres, il y a comparaison des répartitions de l'année en cours et de l'année précédente, ce qui permet de déceler les éventuels problèmes. Si on ne peut corriger ceux-ci au NCHS, on sollicite un fichier de remplacement.

Comme résultat final des opérations de traitement, on obtient des fichiers sous la forme standard avec une même structure de variables, des codes valides ICD-9-CM pour l'année d'enquête et des données d'une qualité acceptable. À l'aide de ces fichiers, on fusionne les données des hôpitaux qui ont fourni un fichier-recensement de leurs congés et on obtient ainsi un fichier unique à des fins d'échantillonnage.

7. TRAITEMENT DES DONNÉES MANUELLES ET AUTOMATISÉES

C'est l'étape où on soumet à des vérifications informatiques poussées tous les dossiers de la NHDS, qu'il s'agisse de données automatisées ou manuelles. Il y a vérification de chaque dossier, c'est-à-dire des contrôles de validité d'étendue des variables non médicales et des contrôles de cohérence des dates d'hospitalisation, de déshospitalisation et d'accouchement. On vérifie une fois de plus les données médicales (diagnostics et actes) en se reportant à une liste de codes valides ICD-9-CM. Les préposés à la vérification font une suite de tests de cohérence, au moyen de tableaux de logique décisionnelle, sur les codes ICD-9-CM selon l'âge et le sexe. Si

le sexe ou l'âge d'un patient est incompatible avec les données médicales d'un dossier, la priorité est accordée aux indications médicales.

On complète les vérifications informatiques par un examen manuel de dossiers choisis qui ont été rejetés par le programme de vérification. Il peut s'agir de codes ICD-9-CM non valides comme mentions initiales ou uniques, d'hospitalisations de plus de 100 jours, d'hospitalisations en obstétrique de plus de 30 jours ou de mentions d'âge de 100 ans et plus. Au NCHS, un nosologiste fait un contrôle manuel de ces dossiers (il y en a de 300 à 500 chaque année) et corrige les incohérences flagrantes.

Il y a imputation de valeurs manquantes pour deux variables, celles de l'âge et du sexe. Ces variables sont manquantes dans moins de 1 % des dossiers. L'imputation d'âge-sexe se fait en fonction de la répartition connue du diagnostic de première mention pour les valeurs connues d'âge et de sexe respectivement. On se trouve ainsi à préserver les répartitions âge-sexe pour ces diagnostics de première mention.

8. CONTRÔLE DE LA QUALITÉ DES DONNÉES MANUELLES ET AUTOMATISÉES

Nous avons conçu une méthode assez étendue de comparaison des estimations NHDS d'une année à l'autre. Si nous l'avons fait, c'est à cause des éléments d'incertitude introduits par le traitement de masses de données électroniques, mais la méthode a aussi éminemment servi à repérer les variations possibles de l'utilisation des services hospitaliers pour un complément d'examen. Un groupe d'analystes est affecté à une tâche que nous appelons le contrôle des données. Cette étape a pour objet d'étudier les écarts entre les estimations de l'année en cours et de l'année précédente. Nous avons élaboré des programmes qui produisent des tableaux à l'intention des analystes de la NHDS. L'opération ne peut être entièrement automatisée, car il faut bien connaître l'enquête, les données sur les patients hospitalisés, les éventuelles tendances des services de santé et les facteurs susceptibles d'influer sur leur prestation.

Voici un bon exemple de la contribution apportée par le contrôle des données à l'amélioration de la qualité de la NHDS. Lorsque nous avons vérifié les données de l'année d'enquête 1993, nous avons constaté une diminution des niveaux et des taux de déshospitalisation d'enfants de moins de 15 ans. Toutes sortes de circonstances peuvent concourir à une évolution des tendances des hospitalisations : refonte du régime d'assurance-maladie, variation d'incidence d'une maladie de l'enfance ou changement de technologie qui fait passer un acte médical au secteur des soins ambulatoires (myringotomie, par exemple). Avant de songer à des causes extérieures, nous devons d'abord établir si la variation numérique en question ne serait pas un effet artificiel de l'enquête. Dans l'exemple cité, l'étude entreprise a révélé que la moitié des hôpitaux spécialisés pour enfants qui étaient visés par l'enquête avaient refusé de participer en 1993. Nous procédons à des corrections de non-réponse d'hôpitaux à l'intérieur des strates, mais pour nous les hôpitaux pour enfants ne sont malheureusement pas une strate d'échantillonnage. Constatant l'effet que cet élément particulier de non-réponse pouvait avoir sur les estimations NHDS relatives aux enfants, nous y sommes allés d'efforts concertés pour remobiliser les hôpitaux en question dans la NHDS de 1994. Depuis, les hôpitaux sélectionnés ont répondu à l'enquête. Nous continuons à suivre ce groupe d'hôpitaux pour y relever tout problème de non-réponse.

9. ORIENTATIONS ET DÉFIS DE L'AVENIR

9.1 Introduction

La mise en application en 1983 aux États-Unis d'un régime de paiement fixe (* prospective payment system + ou PPS) avec des diagnostics regroupés pour la gestion (DRG) à l'intention des patients hospitalisés visés par le régime d'assurance-maladie a suscité des inquiétudes au sujet des effets du programme sur la qualité du codage médical et donc de l'utilité de ces données en recherche. Ainsi, on a craint que le codage médical à des

fins de remboursement soit entaché d'une * déformation DRG +où les gens * surcoderaient +pour accroître les remboursements. Il subsiste des problèmes de fiabilité du codage ICDA, mais la plus grande importance du codage médical pour les remboursements a fait mettre plus l'accent sur cette activité dans les hôpitaux. Il se pourrait même que le codage soit plus sûr que ne l'exige le régime des diagnostics regroupés pour la gestion (Fisher, 1992). L'application du régime des paiements fixes a largement transformé la collecte et l'exploitation des données des dossiers médicaux. C'est dans le contexte des grands changements apportés à la NHDS ou aux programmes nationaux que nous évoquerons plusieurs défis de demain.

9.2 Données pharmaceutiques dans la NHDS

De plus en plus, le NCHS s'est montré désireux d'obtenir plus de données qu'il n'en recueille par ses enquêtes sur les services de santé. C'est ainsi que, en l'an 2000, nous avons rencontré un grand nombre d'organismes fédéraux du domaine de la santé : instituts nationaux de la santé, Département des Anciens combattants, Agency for Healthcare Research and Quality, Food and Drug Administration, Health Resources and Services Administration, Centers for Disease Control and Prevention, etc. Nous avons sondé leur désir d'étendre le champ d'observation de la NHDS au domaine pharmaceutique. Plusieurs thèmes d'intérêt commun pour ces organismes se dégagent, mais ce qui prédominait, c'était le désir de voir recueillir des données pharmaceutiques. L'intérêt pour l'information pharmaceutique s'est aussi manifesté par la création d'un groupe de travail sur les données antimicrobiennes où sont représentés des organismes relevant de tout le Département de la Santé et des Services sociaux.

En 2001, on a affecté des fonds pour déterminer s'il est possible d'effectuer la collecte de données pharmaceutiques pour les patients hospitalisés visés par la NHDS. Le projet est en cours. Jusqu'ici, nous avons fait une première recherche des faits et rencontré des chercheurs de la Maryland School of Pharmacy et des professionnels de l'American Society of Health System Pharmacists. Un de nos buts au départ était de trouver la ou les meilleures sources de telles données pharmaceutiques à relier aux données recueillies dans le cadre de la NHDS. Ces indications serviront à l'établissement des formulaires, des procédures et du matériel de formation initiaux en vue d'un essai sur le terrain.

9.3 * Health Insurance Portability and Accountability Act +

Des mesures législatives récemment adoptées tiendront une grande place dans la réalisation de la NHDS ces prochaines années. Il s'agit de la loi américaine sur la transférabilité des garanties et la responsabilité en matière d'assurance-maladie (* Health Insurance Portability and Accountability Act +ou HIPAA), qui a été promulguée en 1996 et vise à améliorer la transférabilité et la permanence des garanties d'assurance-maladie, à combattre les gaspillages, les fraudes et les abus dans l'assurance-maladie et la prestation de services de santé, à promouvoir l'utilisation de comptes d'épargne-services médicaux et à rendre plus accessibles les services et les assurances de soins de longue durée. Il y a notamment des dispositions de réduction des coûts et des charges administratives des services de santé par une transmission électronique uniforme pour un grand nombre d'opérations administratives et financières, ainsi que des règles de protection des renseignements personnels et de sauvegarde de la confidentialité des données sur les patients. Ce sont ces règles qui, dans la réglementation HIPAA, inquiètent les gens qui travaillent en statistique de la santé. Le degré d'incertitude des règles de protection des données des patients risque d'influer sur les taux de réponse des enquêtes que mène le NCHS auprès des fournisseurs de services de santé. Ce n'est pas en soi une question de contrôle de qualité, mais des taux de réponse moindres nuiront à la fiabilité des estimations de la NHDS.

9.4 Passage de l'ICD-9-CM à l'ICD-10-CM

L'ICD-9-CM demeure la norme de codage de la morbidité aux États-Unis. De concert avec les * Centers for Medicare and Medicaid Services +(que l'on appelait auparavant la * Health Care Financing Administration +), le NCHS participe aux travaux du comité de coordination et de mise à jour de l'ICD. C'est par ce comité que

nous suivons le cours général de la conversion à l'ICD-10 pour la morbidité. Cette dixième version de la nomenclature ne devrait pas s'appliquer avant 2005. Quelle que soit la date de sa mise en application, le passage à la 10^e version exigera d'immenses ressources si nous entendons élaborer ou modifier les systèmes d'information sur les patients hospitalisés pour des organismes utilisateurs comme les hôpitaux, les sociétés d'assurances, les services d'information des États et le gouvernement fédéral. On peut aussi dire que la NHDS sera touchée par ce passage sous presque tous ses aspects. Heureusement, on a prévu un sursis de deux ans entre la notification de la conversion et la mise en application de l'ICD-10. Avec des ressources suffisantes, les responsables de la NHDS auront le temps de se préparer.

10. RÉSUMÉ

Nombreuses sont les études qui ont fait ressortir la nécessité d'améliorer la qualité des données des dossiers médicaux. S'il y a une formation permanente des préposés aux archives médicales des hôpitaux et une bonne communication au sujet des modifications annuelles de l'ICD-9-CM, il n'y a en revanche aucun effort national en vue d'améliorer la qualité des données de base des dossiers médicaux. Au NCHS, nous appliquons des procédures de vérification de la qualité des données de l'enquête NHDS, mais il convient de noter que cette information dépend entièrement de la documentation médicale existante. Le D^r Jencks s'est peut-être le mieux exprimé là-dessus dans un éditorial du *Journal of American Medical Association* (Jencks, 1992) en disant que l'information du dossier médical moyen coûte cher en efforts médicaux, en souffrances des patients et en argent et que nous avons donc l'obligation de veiller à ce qu'elle soit assez bien consignée pour contribuer à l'amélioration des futurs soins aux patients. Si on s'efforçait à l'échelle nationale d'améliorer ces données, tous les systèmes d'information du palier fédéral et des États en profiteraient, tout comme les études rétrospectives qui s'appuient sur les dossiers médicaux.

On peut nettement voir que l'avenir fera évoluer la collecte et l'exploitation des données sur les congés d'hôpital avec des effets sur la NHDS. Les nouvelles exigences de déclaration et de traitement influenceront considérablement sur notre façon de mener cette enquête. Il faudra s'efforcer davantage de mesurer et de garantir la qualité de ces données. Le contrôle de la qualité a toujours tenu une place essentielle dans les opérations de la NHDS, mais ce facteur sera encore plus critique lorsque nous aborderons ces importantes activités nouvelles de demain.

BIBLIOGRAPHIE

- Dennison C, Pokras R. (2000), Design and Operation of the National Hospital Discharge Survey: 1988 Redesign. National Center for Health Statistics. *Vital and Health Statistics* 1(39).
- Duggar BC, et coll. (1981), Alternative Sources for Producing National Statistics on Hospital Utilization. Contract NCHS 223-81-2057, JRB Associates.
- Final Report of the Panel to Evaluate the Cooperative Health Statistics System. (1980), Directions for the '80s. National Center for Health Statistics. Hyattsville, Maryland. (PHS) 80-1204.
- Fisher ES, Whaley FS, Krushat WM, et. coll.. (1992), Accuracy of medicare's hospital claims data: progress has been made, but problems remain. *American Journal of Public Health*, 82, pp. 243-248.
- Green J, Wintfeld N. (1993), How accurate are hospital discharge data for evaluating effectiveness of care? *Medical Care*, 31, pp. 719-731.
- Haupt BJ, Kozak LJ. (1992), Estimates From Two Survey Designs: National Hospital Discharge Survey. National Center for Health Statistics. *Vital and Health Statistics* 13(111).

- Hsia DC, Ahern CA, Ritchie, et coll. (1992), Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988. *Journal of the American Medical Association*, 268, pp. 896.
- Institute of Medicine. (1977), Reliability of Hospital Discharge Abstracts. Washington, D.C. National Academy of Sciences.
- Institute of Medicine. (1980), Reliability of National Hospital Discharge Survey Data. Washington, D.C. National Academy of Sciences.
- Jencks, SF. (1992), Accuracy in Recorded Diagnoses. *Journal of the American Medical Association*, Vol. 267, No. 16, pp. 2238-2239.
- Lissler JT. (1977), Hospital Discharge Survey Evaluation Study. Contract NCHS 255U-1178, Research Triangle Park, North Carolina: Research Triangle Institute.
- McLemore T. (2000). Health Care Establishment Surveys of the National Center for Health Statistics, @ *Proceedings of the Second International Conference on Establishment Surveys*. American Statistical Association. S-11, pp. 1181-1186.
- Miller H., et coll. (1984), Optional Methods for Collecting Data for the National Hospital Discharge Survey. Contract NCHS 282-84-2107, Center for Health Policy Studies.
- National Center for Health Statistics. (1980), International Classification of Diseases, 9th Revision, Clinical Modification. DHHS Pub. No. (PHS) 80-1260. Public Health Service. Washington. U.S. Government Printing Office.
- Simmons WR. (1970), Development of the Design of the NCHS Hospital Discharge Survey. National Center for Health Statistics. *Vital and Health Statistics* 2(39).
- Task Force on Definitions. (1977), The Cooperative Health Statistics System: Its Mission and Program. National Center for Health Statistics. *Vital and Health Statistics* 4(19).