

L'ÉTUDE DE LA QUALITÉ DES DONNÉES DE LA BASE CANADIENNE DE DONNÉES SUR LES CONGÉS DES PATIENTS

Julie Richards, Ann Brown, Craig Homan¹

RÉSUMÉ

La Base de données sur les congés des patients (ou DAD pour *Discharge Abstract Database*) est l'un des dépôts de données importants tenu à jour par l'Institut canadien d'information sur la santé (ICIS). Ce dernier est un organisme national, sans but lucratif, qui joue un rôle essentiel dans l'élaboration du système canadien d'information sur la santé. La DAD contient des données sur les sorties des patients ayant reçu des soins de courte durée transmises à l'Institut par la plupart des hôpitaux canadiens. Les données recueillies sont indispensables à la détermination, par exemple, du nombre et du genre d'interventions et de la durée des hospitalisations. L'ICIS est en train de réaliser la première étude nationale de la qualité de certaines données cliniques et administratives contenues dans la DAD. Cette étude a pour but d'évaluer l'exactitude des données de la DAD par comparaison des données d'origine aux renseignements qui figurent dans la base de données de l'ICIS, afin de repérer toute discordance et d'en établir les causes. Le présent article décrit l'étude de la qualité des données de la DAD et certains résultats provisoires. Ces résultats sont également comparés brièvement à ceux d'une étude similaire. Pour conclure, l'article décrit les prochaines étapes de l'étude et la façon dont les résultats des travaux de la première année permettent d'améliorer la qualité de la DAD.

1. INTRODUCTION

L'Institut canadien d'information sur la santé (ICIS) est en train de réaliser la première étude nationale de l'exactitude des données contenues dans la Base de données sur les congés de patients (ou DAD). L'étude a pour but de repérer les discordances et d'en déterminer les causes, et de fournir aux utilisateurs des résultats de l'étude des renseignements statistiquement fiables sur l'exactitude des données de la DAD. L'étude, d'une durée de trois ans, comportera trois volets annuels. Le rapport décrit d'abord le contexte et les objectifs de l'étude, puis résume la stratégie et la méthodologie et, enfin, présente certains résultats provisoires. Ces résultats sont également comparés à ceux d'une étude similaire. L'article se conclut par une description des prochaines étapes de l'étude.

1.1 Contexte

Fondé en 1994, l'ICIS est un organisme national sans but lucratif qui joue un rôle central dans le développement du système d'information sur la santé au Canada. Il a pour mandat de coordonner le développement et la tenue à jour d'un système global et intégré d'information sur la santé au Canada. Le rôle des dépôts de données de l'ICIS dans le soutien des débats publics et des prises de décision concernant le système de santé au Canada ne cesse de prendre de l'importance.

Afin de remplir son mandat, l'ICIS a mis en place un programme complet et systématique de qualité des données. L'objectif est d'améliorer la qualité des dépôts de données existants et d'assurer que les nouveaux dépôts de données et les produits d'information répondent aux normes de qualité en harmonie avec les objectifs des programmes de l'ICIS et à son engagement à faire preuve d'excellence. Le programme

¹ Julie Richards, Ann Brown et Craig Homan, Institut canadien d'information sur la santé, 377 rue Dalhousie, bureau 200, Ottawa (Ontario), CANADA, K1N 9N8. Ann Brown a été détachée à l'ICIS par Statistique Canada.

d'évaluation de la qualité des données comprend la mise en œuvre d'un cadre de la qualité des données et d'études spéciales se concentrant sur des problèmes de qualité particuliers. Le cadre de référence fournit les critères nécessaires pour l'analyse et la description des dimensions de la qualité des données (et de leurs caractéristiques particulières), à savoir l'exactitude, l'actualité, le caractère utilisable, la pertinence et la comparabilité. Étant donné sa taille, sa couverture et son importance, la première étude spéciale sur la qualité des données a été conçue pour évaluer l'exactitude des données de la Base de données sur les congés des patients ou DAD. Cette étude visait à déterminer la fiabilité d'environ 50 éléments de données figurant sur le résumé analytique des dossiers de sortie de l'hôpital.

1.2 La Base de données sur les congés des patients (ou DAD)

La Base de données sur les congés des patients est l'un des principaux dépôts de données de l'ICIS. Elle contient des données cliniques et administratives ayant trait aux services de santé prodigués aux patients. En se basant sur un ensemble de données normalisées, les hôpitaux préparent un rapport sommaire de sortie (résumé analytique) qui contient des renseignements extraits des dossiers des patients. Cette information est ensuite transmise à l'ICIS où elle est soumise à d'importantes vérifications avant d'être saisie dans la base de données. La population cible de la DAD inclut les sorties de personnes hospitalisées (dans les établissements de soins de courte durée, de soins prolongés et de réadaptation) et les cas de chirurgie d'un jour. Les données transmises à l'ICIS en vue de leur inclusion dans la DAD couvrent environ 75 %² des sorties de l'hôpital, ce qui représente environ 4,3 millions d'enregistrements par an pour les malades hospitalisés et les cas de chirurgie d'un jour. D'après les données sur les diagnostics et sur les interventions qui figurent dans chaque rapport sommaire, l'ICIS calcule des indicateurs de la santé, tels que le taux de cas déclarés de pneumonie et de pontages aorto-coronariens par greffe.

La DAD contient des données démographiques (par exemple, code postal, date de naissance, etc.), des données administratives non médicales (par exemple, numéro d'assurance-maladie, date de l'hospitalisation, etc.), des données cliniques (par exemple, diagnostics, interventions, etc.) et des éléments de données dérivés à valeur ajoutée. Durant l'exercice visé par l'étude (1999-2000), deux systèmes de classification étaient en application pour les diagnostics (c.-à-d. la Classification internationale des maladies, traumatismes et cause de décès, neuvième révision (CIM-9) et la CIM-9 - Modifications cliniques (CIM-9-CM)), ainsi que pour les interventions (c.-à-d. la Classification canadienne des actes diagnostiques, thérapeutiques et chirurgicaux (CCA) et le volume 3 de la CIM-9-CM). À partir de l'exercice 2001-2002, le Canada adoptera les nouvelles classifications des diagnostics et des interventions, à savoir la Classification statistique internationale des maladies et des problèmes de santé connexes, dixième révision Canada – Modifications canadiennes (CIM-10-CA) et la Classification canadienne des interventions en santé (CCI). Parallèlement, l'ICIS met en œuvre un nouveau résumé analytique DAD pour s'adapter à ces nouveaux systèmes nationaux de classification et à l'évolution des besoins de renseignements sur la santé des intervenants. Ce nouveau résumé analytique est conçu de façon à améliorer la normalisation interprovinciale, les couplages entre bases de données et registres, ainsi que certains éléments de données, à ajouter de nouveaux éléments de données et à supprimer les éléments de données non pertinents.

Les données de la DAD sont utilisées par divers organismes à des fins multiples, y compris la surveillance des pratiques d'utilisation des hôpitaux, l'analyse des problèmes de santé et des traumatismes et le suivi de l'état des patients. La DAD est une source importante de données pour la production du rapport annuel de l'ICIS sur les divers indicateurs de la santé qui reflètent les grandes orientations stratégiques et qui sont sanctionnés par la Conférence des sous-ministres de la Santé.

² À l'heure actuelle, les établissements du Québec et certains établissements du Manitoba ne transmettent aucune donnée en vue de leur inclusion dans la DAD.

1.3 Respect de la vie privée, confidentialité et sécurité

L'ICIS a élaboré des politiques très rigoureuses en ce qui concerne le respect de la vie privée, la confidentialité et la sécurité des données. Afin de respecter la vie privée et de sauvegarder la confidentialité des dossiers individuels et des données hospitalières, plusieurs procédures ont été élaborées et suivies durant l'étude. Les employés de l'ICIS ont signé une entente de confidentialité avec les hôpitaux participants et l'ICIS a convenu de ne pas diffuser le nom des hôpitaux participants. Tous les résultats, sauf les rapports transmis aux hôpitaux participants, sont diffusés sous forme agrégée uniquement, de sorte qu'il est impossible d'identifier des patients, des médecins ou des établissements particuliers inclus dans l'étude.

2. CONCEPTION ET MISE EN ŒUVRE DE L'ÉTUDE

2.1 Objectifs de l'étude

L'objectif général de l'étude est d'évaluer l'exactitude de certaines données administratives, au niveau national, provenant de la Base de données sur les congés des patients de l'ICIS. Les objectifs proposés de l'étude, sur une période de trois ans, consistent à évaluer l'exactitude globale de la DAD, l'effet de la collecte de données à partir de dossiers incomplets, la qualité du codage des renseignements sur les diagnostics et les interventions ayant un rapport avec les indicateurs de l'ICIS représentés dans le cadre de référence des indicateurs de la santé et la fréquence à laquelle les renseignements sur les diagnostics et les interventions ne sont pas codés conformément aux lignes directrices de l'ICIS, et à déterminer si des lignes directrices supplémentaires sont nécessaires.

Les éléments de données proposés et les indicateurs de la santé connexes qu'il convient d'évaluer sont précisés pour chaque année de la période d'étude de trois ans. La première année, l'étude s'est concentrée sur les indicateurs de la santé suivants : hospitalisations pour une blessure; accouchement vaginal après césarienne, césariennes, hospitalisations pour une pneumonie ou une grippe, états propices aux soins ambulatoires, pontages aorto-coronariens par greffe et arthroplasties complètes de la hanche.

2.3 Plan de sondage

La population cible comprend tous les séjours de personnes hospitalisées dans les hôpitaux provinciaux de soins de courte durée qui transmettent des données aux fins de leur saisie dans la DAD. La population étudiée comprend environ 2,5 millions de séjours patients (dossiers) dans environ 550 hôpitaux de soins de courte durée.

L'étude a été réalisée par échantillonnage stratifié à plusieurs degrés (Brown et Richards, 2001). Le premier degré correspond à la sélection aléatoire d'hôpitaux. Le deuxième degré correspond à la sélection aléatoire d'un échantillon de dossiers de sortie de l'hôpital stratifié selon l'indicateur de la santé pour chaque établissement participant. On a sélectionné un échantillon représentatif de dossiers de patient au hasard dans chaque strate d'échantillon. En outre, on a sélectionné au hasard des dossiers de patient ne contenant aucun des indicateurs susmentionnés.

En raison de contraintes opérationnelles, l'étude a couvert en tout 18 hôpitaux la première année. Afin d'obtenir un plan d'échantillonnage optimal, on a supposé que, pour chaque indicateur (au niveau national), la proportion de dossiers qui contenaient une discordance était de 15 % et que le degré de fiabilité requis pour l'échantillon correspondait à un coefficient de variation (c.v.) de 16,5% (autrement dit, une erreur-type de 2,5 %). En se fondant sur cette hypothèse et sur l'exigence de fiabilité, on a ensuite établi la taille minimale d'échantillon pour chaque indicateur. Puis, on a augmenté cette taille d'échantillon de 10 % pour tenir compte de la non-réponse (non disponibilité du dossier) et de 10 % supplémentaire pour tenir compte des situations éventuelles de productivité meilleure que prévue de la part des personnes chargées d'extraire de nouveau les données. La taille initiale d'échantillon pour les indicateurs de la santé de l'ICIS était de 2 271 dossiers pour une taille cible globale d'échantillon de 1 950 dossiers.

2.4 Collecte et traitement des données

Les spécialistes de la classification de l'ICIS ont de nouveau extrait les données aux fins de l'étude en retournant à la source originale, pour chaque établissement, pour une période d'une semaine comprise entre septembre et novembre 2000. Aux fins de l'étude, ces spécialistes ont reçu une formation spéciale comportant des renseignements sur l'instrument de collecte, des exercices et des occasions de normaliser des problèmes de codage en suspens.

L'ICIS a conçu et développé une application pour ordinateur portable afin de faciliter la réextraction des données sur place et le dépistage de toute discordance éventuelle entre les données de la DAD et les données réextraites. Toutes les données originales provenant de la DAD ont été téléchargées dans l'application juste avant la semaine de collecte. L'application comportait des listes déroulantes de codes et de causes de discordance, ainsi qu'une zone réservée aux commentaires qui permettait la saisie de renseignements supplémentaires ayant trait à la discordance. Toute documentation supplémentaire qui serait normalement à la disposition des codeurs a également été chargée dans l'ordinateur portable.

Nous avons conçu et développé un système de traitement susceptible d'être réutilisé lors des années suivantes de l'étude. Les fichiers distincts de l'application ont été fusionnés et restructurés aux fins de l'analyse.

Le spécialiste de la classification de l'équipe d'évaluation de la qualité des données a passé en revue manuellement les données cliniques sur les diagnostics et les interventions extraites une nouvelle fois et s'est assuré qu'un code de discordance correct avait été attribué, au besoin. Cet examen a donné lieu à la révision de moins de 2 % des données sur les discordances observées pour les diagnostics et les interventions.

2.5 Dépistage des discordances et de leurs causes

Les éléments de données qui ont fait l'objet d'une nouvelle extraction incluent les caractéristiques démographiques, les renseignements non médicaux, le diagnostic, l'intervention et les produits sanguins. Pour les diagnostics, les interventions et les produits sanguins, toutes les données cliniques ont été réextraites à l'aveugle. Une fois la saisie des nouvelles données achevée, on a repéré les discordances et sélectionné les codes de discordance et de cause. Les données non médicales, comme la date de naissance ou la date de sortie n'ont pas été réextraites, mais ont été comparées aux données originales afin de déterminer s'il s'agissait d'une concordance ou d'une discordance. En cas de discordance, les données non médicales ont été réextraites et les codes de discordance et de cause ont été sélectionnés selon les listes déroulantes. Si le cas nécessitait des éclaircissements, des commentaires supplémentaires ont été inscrits dans la zone réservée à cet effet.

Il y a discordance entre les données lorsque l'élément de données réextrait diffère de l'élément de données original dans la DAD. Le code de discordance indique le genre d'écart entre l'élément de données figurant dans la DAD et l'élément réextrait. Généralement, les discordances étaient du genre suivant :

- données d'entrée manquantes (la personne chargée de réextraire les données a saisi les données non contenues dans la DAD originale);
- données d'entrée non réextraites (la personne chargée de réextraire les données n'a pas codé les données qui figuraient dans la DAD originale);
- données d'entrée différentes (la valeur de l'élément de données figurait dans la DAD originale, mais différait de celle de l'élément réextrait).

Pour les zones de diagnostic et d'intervention, un code supplémentaire de discordance indiquait, avant tout, si la discordance avait trait à la détermination du diagnostic principal ou de l'intervention principale. Pour d'autres codes de diagnostic ou d'intervention, on a utilisé des codes de discordance pour indiquer si la personne chargée de réextraire les données était d'accord ou non avec l'importance de l'effet du diagnostic sur le traitement et(ou) la durée de l'hospitalisation (catégorie de diagnostic), si les diagnostics ont été inscrits dans le dossier avant ou après l'admission et si il y avait concordance entre les codes de diagnostic et d'intervention effectivement choisis pour représenter les problèmes de santé et les interventions.

Des causes ont également été attribuées à chaque discordance. Les causes envisagées étaient les suivantes : erreur de transcription, renseignements incohérents ou contradictoires dans le dossier, interprétation différente de la documentation et spécificité du code non justifiée par la documentation. Notons qu'il était possible d'attribuer plus d'une cause à chaque discordance. Dans certains cas, la cause de la discordance entre les données originales et les données réextraites pourrait être due à la non-disponibilité des données, parce qu'il s'agissait d'un élément de données dont la collecte était facultative ou toute autre raison échappant au contrôle de la personne procédant à la nouvelle extraction des données. Les discordances de ce genre n'ont pas été incluses dans les résultats présentés ici.

2.6 Pondération et estimation

Le poids d'échantillonnage qui indique le nombre de dossiers que représente chaque dossier échantillonné a été calculé en se fondant sur la probabilité de sélection. Il a été utilisé pour estimer la proportion de discordances pour chaque élément de données visé par l'étude.

2.7 Erreur non due à l'échantillonnage et erreur d'échantillonnage

Une quantité considérable de temps et d'énergie a été déployée en vue de mesurer et de réduire au minimum les erreurs non dues à l'échantillonnage, puisque ces dernières peuvent se produire pratiquement à n'importe quelle étape d'une étude. Nous avons mis des méthodes d'assurance de la qualité en œuvre à chaque étape de la collecte et du traitement des données en vue de surveiller la qualité de ces dernières. Ces méthodes incluaient l'utilisation de spécialistes chevronnés pour réextraire les données, l'offre d'une formation spéciale à ces spécialistes en ce qui concerne les procédures de l'étude et des vérifications de la qualité pour s'assurer de la logique du traitement des données (Brown, A. et Richards, J., 2001).

Comme la non-réponse est une source importante d'erreur non due à l'échantillonnage susceptible de biaiser les résultats de l'étude, elle a été surveillée minutieusement pendant toute la durée de la collecte des données. Dans le cas de notre étude, la non-réponse pouvait prendre deux formes. La première était la non-réponse d'un hôpital sélectionné au hasard au premier stade du plan d'échantillonnage. Des 26 établissements choisis au hasard, 22 ont accepté d'emblée de participer et 18 de ceux-ci ont été sélectionnés au hasard. La deuxième catégorie de non-réponse était l'absence du dossier. Le nombre visé de dossiers devant faire l'objet d'une nouvelle extraction des données était de 1 950. De l'échantillon initial de 2 271 dossiers sélectionnés pour produire les indicateurs de la santé de l'ICIS, 1 936 ont fait l'objet d'une nouvelle extraction des données, ce qui donne un taux de réponse au niveau du dossier de 84,5 %. Il convient de souligner que l'on a atteint 99,3 % de la taille cible d'échantillon.

Nous avons appliqué des règles de vérification afin de repérer systématiquement et de corriger les données invalides ou incohérentes. Environ 5 % des codes de discordance attribués aux données administratives non médicales ont été corrigés et moins de 1 % l'ont été dans le cas des données diagnostiques.

Comme les résultats de l'étude se fondent sur un échantillon, les estimations sont sujettes à l'erreur d'échantillonnage. Pour évaluer la fiabilité des estimations des proportions de discordances, nous avons calculé le coefficient de variation. Comme il s'agit de résultats provisoires uniquement, nous n'avons pas calculé ces coefficients pour tous les résultats. Lorsque la deuxième année de l'étude sera achevée, les résultats finals incluront des renseignements sur la fiabilité des estimations.

3. RÉSULTATS PROVISOIRES DE LA PREMIÈRE ANNÉE DE L'ÉTUDE

Nous présentons ici les résultats provisoires sur les discordances observées après la première année de réextraction des données pour les indicateurs de la santé, les caractéristiques démographiques, les données non médicales, les diagnostics et les interventions.

3.1 Résultats pour les indicateurs de la santé

Le tableau 1 montre l'estimation de la proportion de résultats faussement positifs et faussement négatifs. Puisque les indicateurs sont des variables à valeur ajoutée qui sont dérivées d'après les données sur les diagnostics et les interventions, la proportion estimée de résultats faussement positifs ou négatifs est le résultat de discordances ayant trait aux diagnostics ou aux interventions.

Tableau 1 : Pourcentage de cas attribués à un autre indicateur après la réextraction (résultats faussement positifs et faussement négatifs)

Indicateur de la santé	% estimation de faux positifs ¹	% estimation de faux négatifs ²
États propices au traitement ambulatoire (diabète, psychose due à l'alcool/aux drogues, abus de drogues sans dépendance, dépression et hypertension)	10,9	6,5
Césariennes	0,1	0
Pontages aorto-coronariens par greffe	1,1	0
Hospitalisations pour une pneumonie	7,1	12,8
Hospitalisations pour une blessure	5,5	7,1
Arthroplasties complètes de la hanche	0,8	0
Accouchement vaginal après césarienne	0,7	10,5
Non attribué à l'une des situations susmentionnées	1,6	1,4

¹Un résultat est faussement positif lorsque les codes de diagnostic et d'intervention qui figuraient dans le résumé analytique original satisfaisaient les critères d'inclusion dans l'indicateur, mais que les codes obtenus après réextraction des données ne satisfaisaient pas ces critères.

²Un résultat est faussement négatif lorsque les codes de diagnostic et d'intervention figurant dans le résumé analytique original ne satisfaisaient pas les critères d'inclusion dans l'indicateur, mais que les codes obtenus après réextraction les satisfaisaient.

Ce tableau montre que les indicateurs associés à des interventions non urgentes, comme le pontage aorto-coronarien par greffe ou l'arthroplastie de la hanche, et ceux associés à un traitement moins complexe que la césarienne ou l'accouchement vaginal après césarienne, sont codés correctement la plupart du temps. Les diagnostics pour lesquels le protocole de traitement est plus complexe et ceux qui sont moins faciles à définir, comme la pneumonie, les blessures et les états propices au traitement ambulatoire donnent lieu à une plus forte proportion de discordances. Il convient de souligner que la proportion de résultats faussement négatifs pour l'accouchement vaginal après une césarienne est nettement plus forte que la proportion de résultats faussement positifs. Ces données confirment l'hypothèse selon laquelle il est plus probable d'omettre de recueillir les données sur une césarienne antérieure que de saisir des données sur une césarienne qui n'a pas eu lieu.

3.2 Résultats concernant les données démographiques et non médicales

Le tableau 2 donne l'estimation du pourcentage d'erreurs pour les éléments de données démographiques et non médicales. Les éléments de données y sont classés par ordre décroissant l'estimation du pourcentage des discordances totales. Pour les données non médicales, les discordances sont subdivisées en trois catégories, à savoir entrée manquante, entrée non réextraite et entrée différente (telles que définies plus haut). Le tableau 2 montre que, pour la zone «Durée de l'attente au service d'urgence», l'information manquait dans les données transmises originalement pour saisie dans la DAD (5,5 %) ou différait de celle obtenue après réextraction (5,1 %), ce qui donne une l'estimation d'erreur de 10,6 % pour cet élément de données.

Tableau 2 : Données démographiques et non médicales, discordances estimées selon l'élément de données¹

Élément de données	Entrée manquante	Entrée non réextraite	Entrée différente	Total
	% estimation d'erreurs ²	% estimation d'erreurs ²	% estimation d'erreurs ²	% estimation d'erreurs ³
Catégorie d'admission	0,1	0,0	13,8	13,9
Durée de l'attente au service d'urgence	5,5	0,0	5,1	10,6
Heure de sortie	0,0	0,0	9,8	9,8
Code postal	0,0	0,0	9,0	9,0
Code de l'entrée	0,0	0,0	6,5	6,5
Code de réadmission	0,0	0,0	5,0	5,0
Admission par ambulance	3,2	0,5	0,6	4,3
Réadmission non planifiée	3,4	0,0	0,0	3,4
Établissement de destination	0,4	1,0	0,3	1,7
Établissement d'origine	0,7	0,3	0,1	1,1
Poids	0,3	0,1	0,7	1,1

Les éléments de données qui suivent ont produit estimation du taux total d'erreurs <1 % : indicateur d'autotransfusion, transfusion de sang, nombre de jours passés à l'USS, heure de l'admission, albumine, numéro d'unité de l'USS, autres produits sanguins, numéro d'assurance-maladie, code supplémentaire de décès, date de naissance, sortie en vie, globules rouges.

Les éléments de données qui suivent ont produit une estimation du taux total d'erreurs de 0 % : deuxième dossier/numéro d'enregistrement, date d'admission, date des dernières règles, code de décès, date de la sortie, date de naissance estimée, sexe, plasma, plaquettes, province d'émission du numéro d'assurance-maladie.

¹L'estimation de la population (N) est de 2 391 440.

²L'estimation de la proportion pour laquelle cette discordance a été repérée pour cet élément de données.

³L'estimation de la proportion pour laquelle toute discordance a été repérée pour cet élément de données.

Le tableau 3 présente la répartition en pourcentage des causes de ces discordances. Les causes sont classées par ordre décroissant de l'estimation pourcentage. Par exemple, le tableau 3 indique que « la personne qui a extrait les données originales a manqué le renseignement dans le dossier » (code de cause P) est la cause la plus courante (27,6 %) des discordances visant les données non médicales.

Tableau 3 : Données démographiques et non médicales – Estimation des discordances selon le code de cause

Code de cause	% dû à la cause entrée manquante ¹	% dû à la cause entrée non réextraite ²	% dû à la cause entrée différente ³	% dû à la cause toutes les catégories de discordance ⁴
A Erreur de transcription	0,0	0,0	0,7	0,5
B Documentation incomplète	0,0	0,2	0,0	0,0
E Spécificité du code non justifiée	0,0	0,7	0,1	0,1
F Interprétation différente	0,4	16,0	4,6	4,2
K Autre codage de zone grise	0,0	0,0	0,4	0,3
L Renseignements incohérents ou contradictoires	0,0	27,9	1,4	1,8
M Codage contraire aux lignes directrices de l'ICIS	5,4	0,0	10,9	9,7
N Politique de l'hôpital	36,1	15,3	10,4	14,9
O Erreur de codage	0,0	0,0	0,0	0,0
P Renseignements manqués dans le dossier	47,1	1,9	24,3	27,6
Q Erreur mathématique/de dénombrement	0,0	0,0	4,6	3,7
R Téléchargement incorrect	0,1	0,7	27,0	21,8
V Autre	0,6	15,2	2,6	2,5
W Aucune cause apparente	10,2	22,1	13,1	12,9
Total	100,0	100,0	100,0	100,0

¹La population estimée (N) de causes pour les discordances liées à une entrée manquante est 366 390.

²La population estimée (N) de causes pour les discordances liées à une entrée non réextraite est 55 817.

³La population estimée (N) de causes pour les discordances liées à une entrée différente est 1 735 497.

⁴ La population estimée (N) de causes pour les discordances liées à l'ensemble des données non médicales est 2 157 704.

Les tableaux 2 et 3 fournissent des renseignements sur les discordances ayant trait aux données démographiques et non médicales. Selon le tableau 2, des discordances n'apparaissent dans plus de 10 % des enregistrements que pour deux éléments de données et dans 5 % à 10 % des enregistrements, pour trois autres éléments de données. Cependant, pour nombre d'éléments de données, des discordances ont été observées pour moins de 1 % des enregistrements.

Les deux éléments de données pour lesquels des discordances ont été observées pour plus de 10 % des enregistrements sont la catégorie d'admission (13,9 %) et la durée de l'attente au service d'urgence (10,6 %). Viennent ensuite, par ordre décroissant de fréquence, l'heure de la sortie (9,8 %), le code postal (9,0 %) et le code de l'entrée (6,5 %). Nombre des discordances concernant la catégorie d'admission sont dues au fait que certains hôpitaux ont classé tous les patients admis par l'entremise du service d'urgence dans la catégorie « urgence » alors que seuls les patients dont la vie est en danger doivent être classés dans cette catégorie. En outre, la détermination du code approprié d'admission des patientes relevant du service d'obstétrique a posé certaines difficultés.

Il convient de souligner que, selon le tableau 3, 27,6 % de discordances ayant trait à des données non médicales étaient dues au fait que les codeurs ont manqué certains renseignements (code de cause P) qui figuraient dans le dossier. En outre, 21,8 % de l'ensemble des discordances visant les données non médicales se sont produites parce que les données ont été téléchargées incorrectement (code de cause R) – Téléchargement des données sur les admissions, congés et transferts (ADT pour Admission Discharge Transfer) incompatible avec le reste du dossier.

3.3 Résultats concernant les codes de diagnostic

Dans la présente étude, nous nous sommes fondé sur quatre éléments distincts pour comparer les codes de diagnostic, à savoir le préfixe, le code effectif, le suffixe et la catégorie de diagnostic. Dans le tableau 4, nous présentons toutes les discordances possibles pour chacun de ces éléments sous forme de total partiel et de ventilation selon la catégorie. Les pourcentages de discordances observés pour chacune de ces catégories sont : diagnostic principal (DxP) – 13,4 %; comorbidité et complications (CC), problème de classification – 11,0 %; autre discordance concernant les diagnostics – 31,2 % et discordance liée au code proprement dit (qu'il s'agisse de DxP ou de CC) – 6,5 %.

Le tableau 5 montre la distribution en pourcentage des causes des discordances ayant trait au diagnostic principal (DxP) et de toutes les autres discordances ayant trait au diagnostic. Par exemple, le tableau indique que « Interprétation différente de la documentation » (code de cause F) est la cause la plus courante (40,1 %) des discordances ayant trait au diagnostic principal.

Tableau 4 : Estimation des discordances ayant trait au diagnostic

Discordances ayant trait au diagnostic	% d'erreurs estimées par rapport à la population estimée
Discordances pour le diagnostic principal (DxP) ¹	
6 – DxP codé comme étant une autre catégorie dans la DAD	6,3
7-DxP manquant dans la DAD	3,2
11-Dx après l'admission classé comme étant le DxP dans la DAD	0,0
13-DxP secondaire codé comme étant le DxP dans la DAD	0,1
15-Non codé, classé comme étant DxP dans la DAD	3,9
Total des discordances liées au DxP	13,4
Discordances ayant trait à la classification du diagnostic CC²	
8-DxCC codé comme catégorie 3 dans la DAD	2,1
10-Dx avant l'admission codé comme Dx après l'admission dans la DAD	0,2
12-Dx après l'admission codé comme étant Dx avant l'admission dans la DAD	0,6
14-Dx secondaire codé comme un DxCC dans la DAD	8,1
Total des discordances liées à la classification du diagnostic	11,0
Autres discordances liées au diagnostic²	
9-DxCC manquant dans la DAD	14,4
16-Dx non réextrait, classé comme CC dans la DAD	16,6
18- Dx de transfert manquant dans la DAD	0,2
19-Dx non réextrait, classé comme Dx de transfert dans la DAD	0,0
Total des autres discordances liées au diagnostic	31,2
4-Préfixe/suffixe du Dx différent³	0,1
5-Code de Dx différent pour la même affection	6,5

¹Pour DxP, le dénominateur du % d'erreurs est 2 504 600 (population estimée).

²Pour les autres Dx, le dénominateur du % d'erreurs est 2 631 600 (population estimée).

³Pour le total des Dx, le dénominateur du % d'erreurs est 5 870 200 (population estimée).

Tableau 5 : Estimation des discordances ayant trait au diagnostic selon le code de cause

Code de cause	% de la cause pour les discordances de DxP ¹	% de la cause pour les discordances pour les autres Dx ²
A Erreur de transcription	0,0	0,3
B Documentation incomplète	0,0	0,0
D Manque de spécificité du code	0,3	3,1
E Spécificité du code non justifiée	1,5	9,3
F Interprétation différente	40,1	12,2
I Pas d'effet significatif sur le traitement	4,8	24,3
K Autre codage de zone grise	8,5	6,9
L Renseignements incohérents ou contradictoires	8,0	1,8
M Codage contraire aux lignes directrices de l'ICIS	15,9	5,6
N Politique de l'hôpital	0,3	0,4
O Erreur de codage	3,6	8,2
P Renseignements manqués dans le dossier	15,2	25,5
Q Erreur mathématique/de dénombrement	0,0	0,0
R Téléchargement incorrect	0,0	0,0
V Autre	1,6	2,1
W Aucune cause apparente	0,1	0,3
Total	100,0	100,0

¹La population estimée (N) de causes des discordances pour DxP est 363 900.

²La population estimée (N) de causes des discordances liées aux autres Dx est 1 672 400.

En ce qui concerne le diagnostic principal (DxP), la discordance repérée le plus fréquemment est celle où la personne qui a procédé à la nouvelle extraction des données a codé un diagnostic comme étant le DxP alors qu'il avait été codé comme étant une autre catégorie de diagnostic lors de l'extraction originale (code 6).

Pour les autres catégories, la majorité des discordances se répartissent en trois groupes :

- le codeur original a saisi un code que la personne qui a procédé à la nouvelle extraction des données n'a pas jugé significatif (code 16);
- la personne qui a réextrait les données a codé une affection importante que le codeur original n'avait pas codé (code 9);
- la personne qui a réextrait les données et le codeur original ont utilisé un code différent pour représenter la même affection (code 5).

Les causes les plus courantes de ces discordances sont les suivantes :

- la personne qui a réextrait les données n'était pas d'accord que le diagnostic avait un effet important sur le traitement (raison I);
- le codeur original a omis de coder des renseignements qui figuraient dans le dossier (raison P);
- interprétation différente de la documentation (raison F).

3.4 Résultats pour le code d'intervention

Le tableau 6 donne l'estimation de la proportion de discordances ayant trait aux interventions. Les résultats sont ventilés selon l'intervention principale, les autres interventions et les discordances liées aux méthodes d'anesthésie. Par exemple, une discordance a été observée entre les données réextraites et les données originales pour 10 % des interventions principales. Le tableau 7 présente les codes de cause individuelle des discordances ayant trait aux interventions exprimées en pourcentage du total des causes. Par exemple, le tableau 7 indique que le fait que le codeur original ait omis de coder des renseignements figurant dans le dossier (code de cause P) est la cause la plus courante (56,0 %) des discordances ayant trait à l'intervention principale.

Tableau 6 : Estimation des discordances ayant trait aux interventions

Discordances ayant trait aux interventions	% d'erreurs par rapport à la population estimée
Discordances ayant trait à l'intervention principale¹	
22- Intervention principale codée comme une autre intervention	0,3
23- Intervention principale manquante	4,9
25- Intervention non codée, code original comme intervention principale	4,8
Total des discordances ayant trait aux interventions principales	10,0
Discordances ayant trait aux autres interventions²	
24- Autre intervention manquante	10,6
26- Intervention non codée, code original d'autres interventions	12,7
Total des discordances ayant trait aux autres interventions	23,3
21- Codes d'intervention différents³	5,3
Discordances ayant trait à l'anesthésie⁴	
27- Catégorie différente d'anesthésie	6,2
28- Catégorie d'anesthésie manquante	2,1
29- Catégorie d'anesthésie non codée	5,6
Total des discordances ayant trait à l'anesthésie	13,9

¹Pour l'intervention principale, le dénominateur du % d'erreurs est 1 167 000 (population estimée).

²Pour les autres interventions, le dénominateur de % d'erreurs est 1 130 900 (population estimée).

³Pour le total des interventions, le dénominateur de % d'erreurs est 2 300 900 (population estimée).

⁴Pour les méthodes d'anesthésie, le dénominateur de % d'erreurs est 930 600 (population estimée).

Tableau 7 : Estimation des discordances ayant trait aux interventions selon le code de cause

Code de cause	% de la cause pour les discordances ayant trait à l'intervention principale ¹	% de la cause pour les discordances ayant trait aux autres interventions ²	% de la cause pour les discordances ayant trait à l'anesthésie ³
A Erreur de transcription	0,0	0,5	0,0
B Documentation incomplète	0,0	0,0	0,0
D Manque de spécificité du code	0,0	0,9	0,0
E Spécificité du code non justifiée	0,3	5,7	0,3
F Interprétation différente	35,6	20,6	0,8
I Pas d'effet significatif sur le traitement	0,0	1,1	0,0
K Autre codage de zone grise	1,3	7,8	0,7
L Renseignements incohérents ou contradictoires	0,3	4,0	0,9
M Codage contraire aux lignes directrices de l'ICIS	2,1	6,2	38,6
N Politique de l'hôpital	0,1	5,4	10,2
O Erreur de codage	0,5	15,6	0,0
P Renseignements manqués dans le dossier	56,0	28,2	45,9
Q Erreur mathématique/de dénombrement	0,0	0,0	0,0
R Téléchargement incorrect	0,0	0,0	0,0
V Autre	1,2	3,7	0,5
W Aucune cause apparente	2,6	0,3	2,0
Total	100,0	100,0	100,0

¹La population estimée (N) de causes des discordances ayant trait aux interventions principales est 154 000.

²La population estimée (N) de causes des discordances ayant trait aux autres interventions est 502 000.

³La population estimée (N) de causes des discordances ayant trait à l'anesthésie est 130 700.

Le tableau 6 illustre les discordances qui suivent : interventions principales – 10,0 %; autres interventions – 23,3 %; discordances ayant trait au code d'intervention proprement dit - 5,3 % et catégories d'anesthésie – 13,9 %.

Les discordances les plus courantes ayant trait à l'intervention sont les suivantes :

- le codeur original avait codé une intervention que la personne qui a réextrait les données n'a pas codée (code 26);
- la personne qui a réextrait les données a codé une intervention que le codeur original n'avait pas codée (code 24);
- la personne qui a réextrait les données et le codeur original ont utilisé un code différent pour représenter la même intervention (code 21).

Les causes les plus courantes de ces discordances sont les suivantes :

- le codeur original a omis de coder des renseignements qui figuraient dans le dossier (cause P) et interprétation différente de la documentation (cause F) pour les interventions;
- le codeur original a omis de coder des renseignements qui figuraient dans le dossier (cause P) et codage contraire aux lignes directrices de l'ICIS (cause M) pour la méthode d'anesthésie.

4. COMPARAISONS D'ÉTUDES

Bien que nous n'ayons connaissance d'aucune étude basée sur la réextraction des données réalisée au niveau national au Canada, l'Ontario Hospital Association, le ministère de la Santé de l'Ontario et l'Hospital Medical Records Institute (HMRI, un prédécesseur de l'ICIS) ont réalisé en 1991 une étude provinciale fondée sur cette méthode. La Baseline (Re-abstracting) Study a été entreprise en vue de déterminer la qualité des données transmises par l'Ontario au HMRI pour l'exercice 1988-1999 (OHA, MSO, HMRI, avril 1991). Bien que les méthodes d'échantillonnage et de réextraction présentent des similarités, il convient de souligner que l'étude courante de l'ICIS est de portée nationale tandis que la Baseline Study ne portait que sur les données de l'Ontario. En outre, la Baseline Study ne semble pas

aborder les discordances suivantes : (code 11) cas où le Dx après l'admission a été classé comme DxP dans la DAD, (code 13) cas où le Dx secondaire a été codé comme étant le DxP dans la DAD, (code 15) cas où le Dx n'a pas été codé, mais a été classé comme étant le DxP dans la DAD, et (code 25), cas où l'intervention n'a pas été codée, mais où le code original était celui d'intervention principale. En outre, la modification des pratiques de codage, telles que le changement des systèmes de classification en Ontario pour passer de la CIM-9 à la CIM-9-CM, les changements quant aux éléments de données recueillis et l'accès plus facile à l'information et aux services de demande de renseignements sur le codage pourrait limiter la comparabilité de ces études. Par conséquent, toute comparaison doit être effectuée avec prudence. Les résultats des comparaisons qui nous ont permis de repérer des éléments communs sont présentés au tableau 8.

Tableau 8 : Comparaison des résultats des deux études

Élément de données	DAD de l'ICIS 1999-2000 ¹ % de discordances	Ontario 1988-1989 ² % de discordances
Date de l'admission	0,0	0,5
Date de la sortie	0,0	0,9
Numéro d'assurance-maladie	0,2	1,3
Code postal	9,0	7,0
Sexe	0,0	0,8
Poids	1,1	2,5
Date de naissance	0,1	1,5
Code d'entrée	6,5	1,9
Établissement d'origine	1,1	2,2
Établissement de destination	1,7	2,7
Code de sortie en vie	0,1	1,0
Code de décès	0,0	0,0
Code supplémentaire de décès	0,2	0,1
Diagnostic principal	13,4	19,3
Intervention principale	10,0	12,3

¹ Pourcentage de discordances pour la population estimée.

² Pourcentage de discordances pour l'échantillon. L'article présente le pourcentage de concordances. Nous avons soustrait ce pourcentage de 100 pour déterminer le pourcentage de discordances.

5. PROCHAINES ÉTAPES

5.1 Lignes directrices concernant la CIM-10-CA/CCI

L'un des principaux objectifs de l'étude est de faciliter l'élaboration et l'amélioration des lignes directrices concernant le codage en prévision de l'adoption de la CIM-10-CA/CCI comme système de classification au Canada. Nous avons résumé les résultats de l'étude en ce qui concerne les discordances ayant trait au codage des diagnostics et des interventions, et les données sont examinées en profondeur par les spécialistes de la classification de l'ICIS. Cette analyse servira de soutien direct au processus courant d'élaboration de lignes directrices concernant le codage et aux séances de formation offertes par l'ICIS. De la sorte, il sera possible d'accorder l'attention appropriée aux problèmes de codage d'une importance nationale durant l'élaboration des lignes directrices.

5.2 Autres utilisations des résultats en vue d'améliorer la Base de données sur les congés des patients

Bien qu'il soit nécessaire de les analyser plus en profondeur pour bien comprendre les problèmes de qualité des données, les résultats peuvent être utilisés pour définir les priorités en ce qui concerne l'amélioration de

la base de données. Si les taux de discordances observés persistent après la mise en œuvre du nouveau formulaire de résumé analytique élaboré pour la DAD, les éléments de données non médicales pour lesquelles l'estimation du taux de discordances est supérieur à 10 % devraient être examinés. Cet examen pourrait nécessiter la participation des établissements et/ou des fournisseurs de systèmes afin de pouvoir aborder de façon complète certains problèmes de qualité. Les processus de suivi incluent la mise en place de mécanismes appropriés par l'ICIS (comme la révision de la documentation qui figure dans le manuel de la DAD, l'organisation de séances de formation et de téléconférences) en vue d'appuyer les améliorations ayant trait aux problèmes susmentionnés ainsi qu'aux pratiques de codage (telles que mentionnées à la section précédente). Selon les résultats du premier volet de l'étude, les codeurs qui manquent des renseignements figurant dans le dossier (code de cause P) sont la cause d'un nombre important de discordances pour tous les types de données (non médicales et cliniques).

5.3 Deuxième et troisième années

Durant la deuxième année de l'étude, l'expérience tirée des travaux réalisés la première année a permis d'améliorer le logiciel d'application, y compris la validation en direct des codes de diagnostic et d'intervention. Bien que l'approche soit la même, durant la deuxième année, l'étude se concentre sur quatre indicateurs de la santé de l'ICIS, à savoir l'infarctus aigu du myocarde, la fracture de la hanche, l'hystérectomie et l'arthroplastie du genou. Cette deuxième étude porte aussi sur un échantillon de dossiers considérés comme n'ayant pas trait à ces indicateurs. Lorsque ce volet de l'étude sera achevé, les résultats des deux années seront regroupés pour produire des données de référence sur l'exactitude de la DAD avant l'utilisation des nouveaux systèmes de classification et du nouveau formulaire de résumé analytique.

Les travaux réalisés pendant la troisième année de l'étude auront pour objectif principal d'examiner la qualité de la nouvelle DAD, c'est-à-dire d'évaluer la qualité du codage selon la CIM-10-CA/CCI et le nouveau résumé analytique. La comparaison de ces résultats à ceux obtenus les deux premières années donnera une idée de l'efficacité du programme national de formation concernant la CIM-10-CA/CCI et la mise en œuvre des nouvelles normes.

BIBLIOGRAPHIE

- Brown A., et Richards, J., *“Quality Measurements of the Canadian Discharge Database”*, Statistics Sweden International Conference on Quality in Official Statistics, (May 2001).
- Canadian Institute for Health Information (Juin 1, 2000), *Bulletin, “DAD Data Quality Study”*, www.cihi.ca
- Canadian Institute for Health Information (2000), *“Improving Timeliness of the Discharge Abstract Database Data Quality Study”*, www.cihi.ca
- Canadian Institute for Health Information (2001), *“Health Care in Canada”*, www.cihi.ca
- Canadian Institute for Health Information (2001), *“Interim Report for the Data Quality Study of the Discharge Abstract Database: First Year National Findings”*, www.cihi.ca
- Canadian Institute for Health Information (2001), *“Products and Services Catalogue”*, www.cihi.ca
- Ontario Hospital Association, Ontario Ministry of Health, Hospital Medical Records Institute (Avril 1991), *“Report of the Ontario Data Quality Re-abstracting Study”*.