

Le projet d'appariement du Recensement et des fichiers de soins de santé du Manitoba : Composante des ménages privés.

par Christian Houle¹, Jean-Marie Berthelot¹, Pierre David²,
Cam Mustard³, D.Sc., Roos L³, PhD, et M.C. Wolfson⁴, PhD

No. 91

11F1900MPF No. 91

ISBN : 0-660-95240-8

Mars 1996

Les auteurs assument seuls la responsabilité des opinions formulés dans le présent document qui ne représentent pas nécessairement le point de vue de Statistique Canada

Also available in English

¹ Groupe d'analyse et de modélisation de la Santé, Statistique Canada, Ottawa, K1A 0T6

² Division des méthodes d'enquêtes sociales, Statistique Canada, Ottawa, K1A 0T6

³ Manitoba Centre for Health Policy and Evaluation, Department of Community Health Sciences, Faculty of Medicine, University of Manitoba

⁴ Direction des statistiques sociales et des institutions, Statistique Canada, Ottawa, K1A 0T6.

RÉSUMÉ

Introduction: Le contexte économique actuel oblige tous les partenaires des réseaux de soins de santé, qu'ils soient publics ou privés, à identifier les facteurs qui conditionnent l'utilisation des services de soins de santé. Pour améliorer notre compréhension des phénomènes qui sous-tendent ces relations, Statistique Canada et le Manitoba Centre for Health Policy and Evaluation viennent de mettre sur pied une nouvelle base de données. Pour un échantillon représentatif de la province du Manitoba, des microdonnées transversales portant sur le niveau de santé des individus, sur leurs caractéristiques socio-économiques et des données longitudinales détaillées portant sur l'utilisation des services de soins de santé ont été couplées.

Données et méthodes : L'enquête sur la santé et les limitations d'activités de 1986-87, le recensement de 1986 et les dossiers de la Manitoba Health ont été couplés (sans utilisation de noms ou d'adresses) en utilisant le progiciel CANLINK. Dans le cadre du projet pilote 20 000 unités de base ont été sélectionnées selon des techniques d'échantillonnage modernes pour constituer la base de données. Préalablement à l'appariement des fichiers, des consultations ont été tenues afin d'établir un cadre visant à protéger la vie privée et à préserver la confidentialité des données.

Résultat : Un taux de couplage de 74 % a été obtenu pour les ménages privés. Une évaluation de la qualité basée sur la comparaison de noms et d'adresses a permis d'établir que le taux global de concordance parmi les paires appariées est de 95,5 %. Le taux d'appariement de même que le taux de concordance varient selon l'âge et la composition du ménage. Les estimations produites à partir de l'échantillon reflètent bien le profil socio-démographique, la mortalité, l'hospitalisation, les coûts et la consommation de soins de santé des résidents du Manitoba.

Discussion : Le taux de couplage de 74 % s'avère satisfaisant en comparaison du taux de réponses rapporté dans la plupart des enquêtes auprès de la population. En raison de l'excellence du taux de concordance et de la précision des estimations provenant de l'échantillon, cette base de données va permettre d'étudier adéquatement l'association entre les composantes socio-démographiques, la santé et l'utilisation des soins de santé pour la province de Manitoba.

MOTS CLÉS : Appariement probabiliste; confidentialité; échantillonnage; données longitudinales; santé; statut socio-économique.

0.0 INTRODUCTION

Plusieurs études ont clairement démontré qu'il existe un lien entre le statut socio-économique d'un individu et la probabilité de son décès au cours d'une période donnée [1,2,3]. D'autres études ont démontré que la prévalence de certaines maladies varie beaucoup en fonction des caractéristiques socio-économiques du secteur dans lequel un individu demeure [4,5,6]. De plus, quelques enquêtes canadiennes ont déjà fourni des données transversales sur l'état de santé et le statut socio-économique des individus, ainsi qu'une information autodéclarée sur l'utilisation des services de santé (Enquête sociale générale de 1991 [7], Enquête Santé Ontario de 1990 [8], Enquête Santé Québec de 1988 et de 1992 [9], Enquête sur la santé et les limitations d'activités de 1986 et de 1991 [10], Canadian Health and Disability Survey of 1984 [11], Enquête Santé Canada de 1978-79 [12]). Cependant, à notre connaissance, il n'existe pas au Canada de base de données longitudinales qui combine une information exhaustive sur la santé, sur l'utilisation de services de santé et sur le statut socio-économique des individus. Dans un effort particulier pour combler ce besoin en information, Statistique Canada et le Manitoba Centre for Health Policy and Evaluation (MCHPE) ont conjointement mis sur pied un projet pilote visant à évaluer la possibilité de créer une telle base de données à partir de données existantes.

L'objectif premier de ce projet pilote est d'évaluer la faisabilité de réunir les trois sources de données suivantes: le recensement de la population de 1986, l'enquête sur la santé et les limitations d'activités de 1986-1987 (ESLA), et le fichier longitudinal sur les soins de santé de la Manitoba Health (MH). La base de données résultant de cette combinaison permettra aux chercheurs d'explorer de nouvelles avenues en matière de déterminants de la santé. Dans cet article, nous décrivons le jumelage des fichiers, la sélection de l'échantillon pour fins d'analyse et les résultats démontrant la représentativité de la base créée tout en validant les techniques utilisées.

0.1 Confidentialité et droit à la vie privée

Il est primordial lors de la création d'une base de données à partir de données administratives et de données d'enquête, d'assurer la confidentialité des données et de prévenir toute intrusion indue dans la vie privée des individus. Conformément aux politiques des organismes collaborant à ce projet, certaines procédures ont été entreprises avant d'apparier ces ensembles de données. Elles comprennent des consultations avec le Commissaire à la vie privée du Canada, le Faculty Committee on the Use of Human Subjects in Research de l'université du Manitoba, et le Comité sur la confidentialité et la législation de Statistique Canada. De plus, le Comité d'accès et de confidentialité de la Manitoba Health a été informé du projet.

Suite à ces consultations et selon les politiques formelles de Statistique Canada, le ministre responsable de Statistique Canada a autorisé l'appariement tel que proposé : il s'agit d'un projet pilote qui vise à évaluer la faisabilité et l'utilité de l'appariement; le nom et l'adresse des individus ne serviront pas à effectuer l'appariement et ils n'apparaîtront pas non plus sur la base de données; l'appariement sera effectué entièrement dans les locaux de Statistique Canada par des personnes assermentées par la loi sur la statistique; seul un échantillon de 20 000 unités appariées servira aux fins de recherche et d'analyse; enfin, l'accès aux données finales sera étroitement contrôlé selon les dispositions de la loi sur la statistique. Finalement, tous usages reliés au nouveau fichier provenant de l'appariement sont couverts par une entente tripartite entre la University of Manitoba, le ministère de la santé du Manitoba et Statistique Canada.

1.0 DONNÉES

Le questionnaire détaillé (questionnaire 2B) du recensement de la population de 1986 contient de l'information socio-économique détaillée qui comprend des variables telles la composition de la famille, les caractéristiques du logement, le type d'occupation, l'origine ethnique, la langue maternelle, ainsi que plusieurs variables ayant trait au revenu et au niveau d'éducation [13,14]. Ce questionnaire a été livré aux personnes résidant au Manitoba le 3 juin 1986 dans une proportion d'environ un ménage sur cinq. Les autres ménages devaient alors remplir un questionnaire abrégé servant uniquement au dénombrement démographique. Ainsi, le fichier utilisé aux fins d'appariement contient 261 861 enregistrements. Les individus représentés par ces enregistrements peuvent vivre dans deux types de logement: privé ou collectif. Les logements collectifs sont par exemple les hôpitaux, les hospices, les maisons de retraite, les instituts pour handicapés physiques, les orphelinats, les instituts psychiatriques, les hôtels/motels, les camps de travail, les prisons, les colonies Hutterites, les résidences militaires, les regroupements religieux, les résidences d'étudiants et les YMCA. En 1986, il y avait au Manitoba 26 161 personnes vivant dans un logement collectif selon le recensement. Cet article se rapporte principalement à la composante des ménages privés.

L'ESLA de 1986-1987 est une enquête post-censitaire canadienne qui vise à identifier les individus limités par le type ou la quantité d'activités quotidiennes qu'ils peuvent exercer en raison de leur santé. Par post-censitaire, on entend qu'une question du recensement (dans ce cas-ci, la question 20 portant sur les incapacités) sert à définir la base de sondage de la population cible. C'est cette population cible qui est ensuite échantillonnée. Un questionnaire pertinent est alors rempli pour chaque personne échantillonnée. Pour l'ESLA, la population manitobaine vivant dans un ménage privé et ayant des incapacités est étudiée à partir d'un échantillon de 5 480 personnes représentant une population de 150 857 personnes ayant au moins une incapacité. L'ensemble de données ainsi créé contient de l'information sur la santé et les limitations fonctionnelles des individus ainsi que sur le type d'emploi, le niveau d'éducation, le transport, le logement et les loisirs. L'enquête étant de type auto-déclaré, les données représentent la situation des répondants de leur point de vue plutôt que d'un point de vue administratif ou clinique.

Le fichier longitudinal de la MH, quant à lui, contient de l'information sur les visites à l'hôpital, les diagnostics, les interventions chirurgicales, les soins de santé reçus à la maison, la date et la cause des décès, ainsi que d'autres données sur l'utilisation des soins de santé. De nombreuses études innovatrices en recherche sur les soins de santé ont utilisé ce fichier (par exemple, [15,16,17]). La date d'entrée en vigueur de la couverture d'assurance maladie et celle d'annulation de la couverture ont servi à créer un registre des personnes couvertes par l'assurance maladie du Manitoba en juin 1986. Ce registre contenait 1 047 443 enregistrements.

2.0 MÉTHODES

Le projet d'appariement comporte trois étapes principales. La première étape consiste à assortir par paires les individus appartenant à trois sources de données distinctes. La seconde étape est une extension naturelle de la première car il s'agit d'évaluer dans quelles proportions les paires formées représentent le même individu. La troisième étape consiste à choisir un échantillon de 20 000 unités appariées afin de créer la base de données pour fins d'analyse. Dans cette section, nous traitons tour à tour de la méthodologie employée à chacune de ces étapes.

2.1 Appariement

Le système Canlink [18,19] développé à Statistique Canada a été utilisé pour l'étape d'assortiment. Il s'agit d'un logiciel d'appariement probabiliste qui assortit les enregistrements de deux ensembles de données en utilisant la puissance discriminante des variables communes disponibles. Le système pondère les paires d'enregistrements en fonction du degré de concordance des valeurs observées et tient aussi compte de la probabilité des concordances aléatoires. Les fichiers à assortir sont celui de l'échantillon 2B du recensement de 1986 couvrant la province du Manitoba et le registre des personnes inscrites à la MH en juin 1986 ne contenant qu'un sous-ensemble des variables disponibles. Seuls ces deux fichiers sont en jeu lors de l'appariement probabiliste puisque l'échantillon de l'ESLA de 1986-1987 ayant été tiré de l'échantillon 2B du recensement [20], tous les enregistrements⁵ de l'ESLA sont déjà assortis à ceux du recensement par une clé unique.

Les enregistrements individuels à assortir proviennent donc de deux fichiers: l'un contenant les enregistrements de 261 861 individus vivant au Manitoba (dérivé du fichier 2B du recensement de 1986) et l'autre contenant les enregistrements de 1 047 443 personnes (un dérivé du registre administratif du Manitoba). La stratégie adoptée pour identifier les paires représentant le même individu (bonnes paires) consiste à morceler les deux ensembles de données en blocs et à ne former que les paires d'individus appartenant au même bloc.

Les paires d'enregistrements ne sont comparées que si toutes les variables de blocage concordent a priori. Un choix judicieux s'imposait donc pour ne pas éliminer, dès cette étape, un grand nombre de bonnes paires. Or, nous le rappelons, les variables les plus discriminantes que sont les noms, prénoms et adresses n'ont pas été utilisées dans cette étude. Cette contrainte nous a obligés à choisir d'autres combinaisons de variables possédant un pouvoir de discrimination plus limité et à renforcer ces choix par l'application de techniques innovatrices.

Deux phases d'appariement ont été exécutées. Dans un premier temps et après avoir examiné diverses définitions possibles, nous avons défini un bloc comme étant un ensemble de quatre caractéristiques individuelles, soit le sexe, l'année de naissance, le mois de naissance et le code postal d'une personne. Lors de la phase II de l'appariement, la définition a été relâchée afin de former davantage de paires d'individus. L'année et le mois de naissance exacts ont été remplacés par l'âge de la personne, ce qui permet de comparer un individu à un plus grand nombre de candidats. De plus, l'aire couverte par la variable géographique en milieu urbain a été agrandie par un facteur d'environ trois puisque le secteur de dénombrement du recensement a remplacé le code postal.

⁵ Sauf pour les ménages collectifs de type institutionnels.

Ces appariements ont permis de diviser le fichier du recensement en trois sous-ensembles : les enregistrements ayant apparié clairement (définitifs), ceux ayant apparié mais dont la puissance discriminante des variables disponibles soulève un doute selon les critères de CANLINK (possibles) et ceux n'ayant jamais apparié.

Bien que l'information sur la structure familiale ait été utilisée lors de l'appariement, le système CANLINK ne compare que deux individus à la fois sans tenir compte des appariements obtenus parmi les autres membres de la famille. Il nous a fallu définir une série de règles afin de s'assurer de la cohérence des appariements à l'intérieur d'une famille et entre les deux phases d'appariement. Un document détaillé concernant cette étape de la méthodologie a déjà fait l'objet d'une publication [21].

2.2 Évaluation de la concordance des paires formées.

Notre évaluation visait deux objectifs. Dans un premier temps, il était important de connaître le degré d'exactitude avec lequel nous avons associé les données du Manitoba et celle du recensement (appariements définitifs uniquement). Ensuite, il fallait évaluer si les règles développées pour rejeter certains appariements «possibles» étaient adéquates.

Un échantillon de 1,000 familles a été tiré représentant 2 102 individus appariés. La région rurale ou urbaine, la composition de la famille (personne vivant seule, couple avec enfant, couple sans enfant, famille multiple) et le statut de l'appariement (définitif ou possible) ont servi de variables de stratification. Le MH a ensuite procédé à la saisie des noms et adresses de tous ces individus ainsi que des autres membres de leur famille. Il est entendu que cette information nominative n'a pas servi à déterminer la validité d'appariements spécifiques, mais seulement à estimer les taux d'appariement véritable à des niveaux agrégés. Ces noms et adresses ont été comparés manuellement aux questionnaires 2B conservés à Statistique Canada sous forme de microfilms.

Seulement 17 cas n'ont pu être comparés, soit à cause d'un microfilm illisible ou parce que les noms et adresses n'avaient pas été retrouvés sur le fichier du MH. Le poids de ces unités a été redistribué à l'intérieur des strates où la non-réponse était identifiée.

2.3 Sélection de l'échantillon.

Comme le projet met en jeu trois bases de données, il faut préciser que notre référence échantillonnale (base de sondage pour l'échantillonnage) est le fichier du recensement 2B de 1986. Avant même d'élaborer un plan d'échantillonnage, plusieurs contraintes s'étaient déjà imposées d'elles-mêmes. Rappelons en effet que la taille de l'échantillon était déjà fixée à un maximum de 20 000 unités et que la base formée devait réunir de l'information des fichiers du recensement, de celui de MH mais aussi de l'ESLA. Or, l'échantillon de l'ESLA utilise l'individu comme unité d'échantillonnage alors que notre analyse de la population globale du Manitoba devrait utiliser le ménage tel que défini par le recensement. Plusieurs options ont alors été envisagées pour tenter de construire une seule base mais la complexité de l'analyse qui en aurait résulté annihilait pratiquement tous les gains en précision. Pour assurer un équilibre entre la simplicité d'analyse et un plan performant, la sélection consisterait à construire deux bases indépendantes: la première pour étudier le lien entre l'incapacité, le statut socio-économique et la santé et la seconde pour analyser la population générale du Manitoba. Pour maximiser l'utilisation des 20 000 unités, il fallait de plus tenir compte du chevauchement de ces deux bases.

Dû au plan d'échantillonnage complexe de l'ESLA, au nombre relativement petit d'individus échantillonnés pour l'ESLA et à l'importance de cette base du point de vue analytique, les individus avec incapacité appariés sont tous retenus. Ceci représente 4 434 unités de base. Cet échantillon forme la première base servant à l'analyse des personnes ayant au moins une incapacité fonctionnelle. La pondération utilisée lors de l'analyse sera celle de l'ESLA ajustée pour tenir compte du taux d'appariement spécifique à certains sous-groupes.

Il restait donc 15 566 unités libres pour former la base globale du Manitoba plus le nombre espéré d'unités chevauchant les deux bases. Poursuivant toujours l'objectif d'optimiser le plan d'échantillonnage, la stratification semblait appropriée. En effet, la stratification est un outil particulièrement performant pour trois dimensions d'une enquête par échantillonnage. Premièrement, elle permet de réduire la variance globale des estimations. Deuxièmement, elle assure un standard de qualité aux estimations pour des sous-groupes d'intérêt de la population. Troisièmement, la stratification peut améliorer la précision dans certains cas où l'on peut tenir compte de l'erreur non-échantillonnale. Finalement, la stratification est d'autant plus efficace lorsque les variables de stratification sont corrélées avec la variable étudiée.

Puisque plusieurs études ont établi des liens entre le statut socio-économique et la santé, il est naturel d'utiliser des variables socio-économiques pour construire les strates. De plus, il n'y a aucun inconvénient à utiliser le ménage comme unité d'échantillonnage puisque le statut socio-économique est globalement le même pour tous les membres.

Rappelons que l'échantillon doit représenter la population manitobaine d'une façon générale. Puisque c'est le fichier du recensement de 1986 qui détermine entièrement la composition de cette population, toutes les variables de stratification⁶ y ont été directement puisées ou en ont été dérivées. La liste suivante décrit comment les cellules de stratification ont été formées :

1. Type de secteur de dénombrement;
2. Structure familial du ménage;
3. Ménage contenant au moins un oui à la question 20;
4. Groupe d'âge de la personne représentant le ménage;
5. Sexe de la personne représentant le groupe d'âge;
6. Région urbaine ou rurale;
7. Tercile de l'éducation de la personne représentant le ménage;
8. Tercile du revenu ajusté pour la composition du ménage.

Pour construire certaines de ces variables, il a fallu définir une personne représentant le ménage. Lorsque des enfants étaient présents, et vu la corrélation entre l'âge de la mère et celle des enfants, il nous a semblé naturel de choisir la femme adulte la plus âgée d'une famille pour la représenter. Autrement, notre choix s'arrêtait sur l'adulte le plus âgé. Le choix de ces variables constitue un judicieux équilibre entre l'uniformisation des taux d'appariement à l'intérieur d'une strate et l'utilisation des variables socio-économiques.

Bien qu'aucun suréchantillonnage n'ait été opéré a priori, quelques regroupements ont ensuite été nécessaires pour respecter des critères de taille minimale d'unités échantillonnales par strate. Le nombre final de strates pour les ménages privés est de 611. Le nombre total d'unités tirées est 16 387. Cela représente 46 670 personnes.

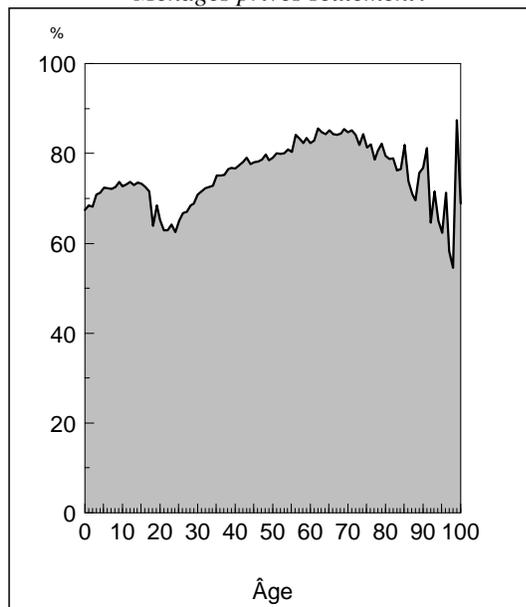
⁶ Le détail de ces variables est présenté en annexe.

Finally, it is common to adjust the sampling weights so that the estimated totals from the sample reproduce as closely as possible the population totals. Poststratification allows adjustment of the weights for categories where the number of units is insufficient to create a true stratum but where the analytical importance justifies the use of special techniques. These techniques modify the initial weights under constraints of minimum change [32]. For private households, the weights were adjusted by age group, by rural or urban geographical region, by marital status and by sex at the individual level, whereas at the household level, the rural or urban geographical region, household size and the status of owner or tenant played the same role for adjustment.

3.0 RÉSULTATS

3.1 Appariement

Figure 1. Taux d'appariement par âge.
Ménages privés seulement.



Malgré toutes les restrictions et les exclusions appliquées aux appariements initiaux, globalement, 74 % (174 476 sur 235 700) des individus du recensement vivant dans un ménage privé ont été appariés à un individu sur le fichier du Manitoba. Ce taux varie selon la mobilité géographique, l'âge, l'état matrimonial et la taille de la famille.

Les facteurs qui ont eu le plus d'influence sur le taux d'appariement sont tous reliés à la mobilité géographique des individus. Ainsi, les groupes d'individus suivants ont été plus difficiles à appairer : les jeunes adultes (entre 20 et 25 ans : figure 1), les personnes ayant changé de lieu de résidence entre les recensements de 1981 et 1986 (tableau 1), ainsi que les personnes divorcées ou séparées (tableau 2). Chez ces groupes, les changements fréquents d'adresse et de structure familiale rendent la concordance entre les deux sources de données plus difficile que chez les groupes moins mobiles. En effet, comme les données du recensement sont en date du 3 juin 1986 précisément et que certaines données du MH sont en date du 31 décembre 1986, un décalage de l'information est plus probable chez les individus mobiles.

Tableau 1. Taux d'appariement selon la mobilité
Ménages privés seulement.

Mobilité	Taux d'appariement %
Même ménage	81,7
Même DR	65,8
Autre DR	62,5

DR : division de recensement, une aire géographique utilisée par le recensement. Le territoire du Manitoba compte vingt-trois divisions de recensement.

Tableau 2. Taux d'appariement selon l'état matrimonial
Ménages privés seulement.

État matrimonial	Taux d'appariement %
Marié	78,5
Veuf	74,5
Célibataire	71,2
Divorcé	61,4
Séparé	43,4

On observe un faible taux d'appariement chez les personnes séparées. Ceci s'explique par la mobilité inhérente au phénomène de la séparation et aussi par le décalage de l'information entre les deux sources de données.

L'effet de l'âge sur le taux d'appariement n'est pas surprenant. Les enfants de moins de quinze ans et les adultes entre trente et soixante ans connaissent de meilleurs taux étant donné leur situation plus stable. Chez les individus de plus de 85 ans, on trouve davantage de variabilité dans les données à cause du phénomène de l'entrée en institution et du petit nombre de cas.

Chez les individus qui n'ont pas déménagé entre les recensements de 1981 et 1986 (même ménage), on aurait pu s'attendre à un taux d'appariement encore meilleur. Le taux de 81,7 % est peut-être une indication qu'il y a un plafond d'appariement d'environ 80 % en utilisant la méthodologie décrite jusqu'ici, étant donné que les fichiers ne sont pas absolument exempts d'erreurs.

Intuitivement, la taille de la famille est corrélée de deux façons opposées avec le taux d'appariement. Bien qu'une famille nombreuse ait une contrainte intrinsèque face à la mobilité du noyau familial, plusieurs membres en revanche viendront s'accrocher ou quitteront ce noyau régulièrement. On peut observer au tableau 3 qu'en effet le taux d'appariement diminue significativement lorsque la taille de la famille augmente.

Tableau 3. Taux d'appariement selon la taille de la famille
Ménages privés seulement.

Taille de la famille	Taux d'appariement %
1	66,4
2	78,5
3	74,7
4	79,8
5	77,1
6	70,0
7	55,9
8	51,4
9	39,2
10+	46,7

3.2 Évaluation de la concordance des paires formées.

Le tableau 4 est rassurant puisqu'il démontre que, globalement, plus de 95 % des appariements définitifs conservés représentent le même individu. L'échantillon de 20 000 unités étant tiré uniquement à partir d'appariements définitifs, il en résulte une qualité exceptionnelle. Dans le même ordre d'idée, le taux de concordance des noms de seulement 40 % parmi les paires possibles indique que nos règles sévères étaient justifiées et judicieuses puisqu'elles ont évité d'accepter de mauvais appariements dans une grande proportion.

La taille du ménage semble être étroitement reliée au taux de concordance. Les personnes vivant seule et celles vivant dans un ménage de huit personnes ou plus connaissent un taux plus bas avec respectivement

86,8 % et 90,8 % de concordance. Le petit nombre de variables discriminantes disponibles pour les personnes seules et la présence de plusieurs familles dans un même ménage pour les ménages de grande taille expliqueraient ces résultats.

Tableau 4. Taux de concordance des noms selon divers regroupements.
Ménages privés seulement.

	Concordance sur les noms %	Erreur-type* %
Appariement «possible»	40,1	2,4
Appariement «définitif»	95,5	0,5
Réserve Indienne	95,3	1,5
Ménage de taille 1	86,8	2,5
Ménage de taille 2 à 7	96,5	0,5
Ménage de taille 8 ou plus	90,8	2,0

* *L'effet de plan est ignoré dans le calcul de l'erreur-type.*

Finalement, on peut observer que les habitants des réserves indiennes appariés⁷ ont un taux de concordance équivalent aux personnes vivant hors réserve.

3.3 Sélection de l'échantillon.

Globalement, nous avons construit la stratification en cherchant à atteindre surtout deux objectifs précis. Premièrement, il fallait s'approcher le plus possible d'un plan auto-pondéré pour permettre l'utilisation d'outils informatiques déjà existant. Les coûts et le temps de développement d'outils spécifiques auraient été un handicap majeur pour toute analyse subséquente. Cet objectif a été atteint en évitant le suréchantillonnage des strates en autant que ce fut possible et en assurant une certaine uniformité des poids à l'intérieur de chaque strate formée.

L'utilisation de variables socio-économiques dans le processus de formation des strates constituait notre second objectif. Quand les tailles le permettaient, les variables dérivées du revenu, de la scolarité, de la structure de la famille, de l'âge et de la géographie ont donc été utilisées.

La formation d'une nouvelle base de données à partir de fichiers administratifs provenant de différentes sources conduit naturellement notre évaluation à comparer les estimations produites par notre échantillon à celles provenant des sources originales. Les tableaux 5 et 6 nous indiquent avec quelle précision nous arrivons à reproduire les comptes du recensement 2B de 1986 après poststratification.

Précisons qu'avant la poststratification, quatre catégories s'éloignaient suffisamment de la cible pour justifier à elles seules une action immédiate. Ce sont : les hommes et les femmes divorcés (écart de +7,8 %), séparés (écart de -17,6 %) ou veufs (écart de +6,1 %) ainsi que les femmes de 0 à 4 ans (écart de -4,6 %). Ces catégories représentent un petit nombre d'individus chacune et il a été impossible de les

⁷ Il faut cependant tenir compte que le taux d'appariement dans les réserves indiennes n'est que de 44,5 %. C'est nettement inférieur au taux moyen de 74 % ce qui pourrait conduire à un biais, les individus appariés ayant des caractéristiques très différentes de la population vivant en réserve indienne.

traiter séparément lors de la stratification à la sélection de l'échantillon. La poststratification a permis de ramener tous les écarts en deçà de 1,5 %.

Tableau 5. Précision de l'échantillon par groupes d'âge.

Ménages privés seulement.

	Hommes Recensement 2B	Différence relative de l'échantillon %	Femmes Recensement 2B	Différence relative de l'échantillon %	Total Recensement 2B	Différence relative de l'échantillon %
0 à 4 ans	39 929	-0,05	37 709	-0,06	77 638	-0,05
5 à 14 ans	78 855	-0,13	76 074	-0,15	154 929	-0,14
15 à 24 ans	84 897	-0,20	84 250	-0,12	169 147	-0,16
25 à 44 ans	155 160	-0,11	157 430	-0,05	312 590	-0,08
45 à 64 ans	93 145	-0,06	98 851	-0,05	191 996	-0,05
65 ans et +	52 829	-0,01	68 309	-0,02	121 138	-0,02
Total	504 815	-0,10	522 623	-0,07	1 027 438	-0,09

Tableau 6. Précision de l'échantillon par état matrimonial.

Ménages privés seulement.

	Hommes Recensement 2B	Différence relative avec l'échantillon %	Femmes Recensement 2B	Différence relative avec l'échantillon %	Total Recensement 2B	Différence relative avec l'échantillon %
Déjà marié	26 649	-0,51	69 037	-0,10	95 686	-0,21
Marié	245 443	-0,03	245 799	-0,03	491 242	-0,03
Célibataire	232 723	-0,13	207 787	-0,12	440 510	-0,12
Total	504 815	-0,10	522 623	-0,07	1 027 438	-0,09
Divorcé	9 146	-0,08	14 388	-0,15	23 534	-0,12
Séparé	9 137	-1,40	11 242	-0,04	20 379	-0,65
Veuf	8 366	-0,01	43 407	-0,09	51 773	-0,08

Finale­ment, il existe des différences conceptuelles importantes au niveau de la définition des populations représentées par le recensement et par celle du fichier du MH. Seules les personnes ayant un domicile habituel au Manitoba le 3 juin 1986 sont recensées dans cette province. Les opérations du recensement ne réussissent pas à faire remplir un questionnaire à toutes les personnes qui ont la qualité de résident ce qui entraîne un sous-dénombrement. D'autre part, le fichier du MH contient les personnes qui possèdent la qualité de couverture du régime d'assurance maladie et d'hospitalisation. Certaines de ces personnes n'habitant plus le Manitoba peuvent parfois encore bénéficier de cette protection ou elles peuvent ne pas avoir indiqué un changement de leur statut, entraînant un certain surdénombrement. Pour plusieurs catégories de résidents collectifs dont les services médicaux sont assurés par le gouvernement fédéral, comme les camps militaires et certaines réserves indiennes, le fichier du MH ne contient aucune information alors que celui du recensement considère ces personnes comme des résidents du Manitoba.

Pour des fins de comparaison, nous avons exclu des comptes du MH des prochains tableaux les personnes vivant dans une résidence pour personnes âgées puisqu'elles résident alors dans un logement collectif institutionnel. De plus, selon la définition du recensement, les personnes ayant séjourné 180 jours ou plus

dans une institution de soins de santé possèdent la qualité de résident permanent. Puisque ces personnes résident donc dans un logement collectif institutionnel, nous les avons aussi exclues des comptes des prochains tableaux. Malgré ces efforts pour uniformiser les populations, il demeure que nous n'avons pu qu'approximer les comptes des personnes vivant en institution. Par complémentarité, les populations comparées représentent approximativement les personnes vivant dans un ménage privé ou dans un ménage collectif non-institutionnel.

Comme le démontre le tableau 7, et malgré les différences conceptuelles importantes, les tailles de ces deux populations par groupe d'âge sont très favorablement comparables l'une à l'autre. Globalement, les tailles totales estimées des deux populations ne diffèrent que par 0,1 %, bien qu'il y ait sous-estimation des hommes par 1,2 % et une surestimation des femmes par 1,4 %. On observe aussi que les plus grandes différences se trouvent parmi les individus les plus jeunes.

Tableau 7. Précision de l'échantillon par groupes d'âge versus MH.
Ménages privés et collectifs non-institutionnels.

	Hommes MH	Différence relative avec l'échantillon %	Femmes MH	Différence relative avec l'échantillon %	Total MH	Différence relative avec l'échantillon %
0 à 4 ans	32 743	2,57	31 105	1,98	63 848	2,28
5 à 14 ans	78 076	1,47	73 912	2,87	151 988	2,15
15 à 24 ans	86 722	-1,24	82 971	1,61	169 693	0,15
25 à 44 ans	165 783	-2,84	159 458	1,92	325 241	-0,51
45 à 64 ans	96 989	-1,71	98 997	0,92	195 986	-0,38
65 ans et +	57 904	-1,34	74 129	-1,10	132 033	-1,20
Total	518 217	-1,20	520 572	1,39	1 038 789	0,10

Les tableaux 8, 9 et 10 permettent de comparer le taux de mortalité, l'utilisation des soins médicaux et l'utilisation des soins hospitaliers selon qu'ils sont estimés à partir de notre échantillon ou du fichier du MH. Notons que les taux de décès rapportés dans la littérature [en particulier [22]] sont légèrement plus haut que ceux présentés au tableau 8, l'écart s'accroissant avec l'âge. Ceci s'explique par l'exclusion de nos fichiers des individus vivant dans un logement collectif institutionnel qui connaissent un taux de mortalité plus important que les personnes vivant en ménage privé.

Tableau 8 Taux de mortalité annualisés et basés sur la période juin 1986 à mai 1989
Ménages privés et collectifs non-institutionnels

Âge	Taux de mortalité annuel* (x 1 000) MH	Taux de mortalité annuel (x 1 000) Échantillon	Intervalle de confiance à 95 % pour l'échantillon
0-4	0,51	0,02	[0 – 0,18]
5-9	0,16	0,20	[0 – 0,66]
10-14	0,24	0,27	[0 – 0,80]
15-19	0,78	0,60	[0 – 1,39]
20-24	0,94	0,70	[0 – 1,52]
25-29	0,83	0,40	[0 – 1,02]
30-34	0,86	0,44	[0 – 1,11]
35-39	1,12	1,07	[0 – 2,19]
40-44	1,85	2,07	[0,37 – 3,77]
45-49	3,23	2,78	[0,57 – 4,99]
50-54	5,12	3,68	[1,03 – 6,33]
55-59	8,42	8,91	[4,80 – 13,02]
60-64	12,43	10,48	[6,01 – 14,95]
65-69	18,96	19,03	[12,69 – 25,37]
70-74	28,63	24,59	[16,76 – 32,42]
75-79	42,77	43,09	[30,85 – 55,33]
80 et plus	76,98	73,67	[57,41 – 89,93]
Total	6,71	6,11	[5,39 – 6,83]

*Nombre de décès estimé (pour la période de trois ans) divisé par 3 fois le total de la population .

Globalement, le taux de mortalité estimé par notre échantillon (6,11) est inférieur à celui dérivé du fichier du MH (6,71), mais cette différence n'est pas statistiquement significative à un niveau de confiance de 95 %. Remarquons que les intervalles de confiance dérivés de notre échantillon contiennent la valeur calculée par le MH pour tous les groupes d'âge sauf les enfants de 0 à 4 ans. Bien que le nombre de décès dans cette catégorie soit relativement petit, il semble que la difficulté à apparier les enfants de moins d'un an soit reliée à cette sous-estimation. Cela indique aussi que toute analyse spécifique aux enfants de 0 à 4 ans devra être menée avec prudence particulièrement lorsque la prévalence d'une maladie ou d'un état est faible.

Tableau 9 Coûts et services médicaux, année fiscale 1986-87
Ménages privés et collectifs non-institutionnels

Type de pratique	Nombre de services			Coûts des services (\$)		
	MH	Échantillon	Différence relative (%)	MH	Échantillon	Différence relative (%)
Médecine Interne	699 542	702 735	0,46	19 060 658	18 904 922	-0,82
Pédiatrie	416 157	449 122	7,92	6 932 217	7 359 290	6,16
Psychiatrie	154 279	146 704	-4,91	8 970 489	8 468 584	-5,60
Chirurgie	406 907	409 097	0,54	21 772 057	21 230 743	-2,49
Ophthalmologie	339 334	357 273	5,29	10 017 371	10 506 461	4,88
Oto-rhino-laryngologie	100 859	110 164	9,23	2 952 711	3 181 710	7,76
Dermatologie	115 516	126 086	9,15	1 955 985	2 114 772	8,12
Radiologie	623 712	653 850	4,83	9 564 330	9 970 191	4,24
Pathologie	2 941 244	3 126 365	6,29	21 369 502	22 489 399	5,24
Obstétrique et gynécologie	294 288	328 728	11,70	8 774 785	9 151 231	4,29
Anesthésie	74 624	74 442	-0,24	5 762 857	5 572 964	-3,30
Pratique générale	4 762 316	4 858 641	2,02	75 806 649	76 545 594	0,97
Médecine physique*	9 325	8 333	-10,64	447 187	413 127	-7,62
Totaux	10 938 103	11 351 540	3,78	193 386 798	195 908 988	1,30

* Traduction libre

Pour la plupart des catégories de type de pratique médicale, les estimations tirées de l'échantillon s'approchent sensiblement des chiffres présentés par le MH, autant pour le nombre de services que pour les coûts engendrés pour produire ces services. La précision atteinte est d'autant remarquable qu'aucune poststratification n'a été faite pour ajuster la consommation de soins de santé au compte du MH et ce à quelque niveau que ce soit.

Tableau 10. Nombre et durée des séjours à l'hôpital, année fiscale 1986-87
Ménages privés et collectifs non-institutionnels

Groupes d'âge	Nombre de séjour			Durée des séjours		
	MH	Échantillon	Différence relative (%)	MH	Échantillon	Différence relative (%)
0-64 ans	100 127	96 303	-3,82	538 616	499 665	-7,23
65 ans et +	43 226	41 318	-4,41	452 172	414 555	-8,32
Total	143 353	137 621	-4,00	990 788	914 220	-7,73

Le tableau 10 révèle les résultats de la comparaison du nombre et des durées des séjours à l'hôpital. Toujours compte tenu des différences conceptuelles entre les sources de données, une précision des estimations à l'intérieur de 10 % est satisfaisante. Une sous-estimation plus importante pour la durée des séjours que pour leur nombre indique que les séjours les plus long sont plus sous-estimés que les séjours de courte durée. La difficulté à identifier les résidents de logements collectifs institutionnels à partir des fichiers du MH explique cet état des choses.

4 0 DISCUSSION

La méthodologie présentée dans cet article permet d'apparier environ 74 % de la partie du fichier du recensement correspondant aux ménages privés au fichier du MH en utilisant principalement l'âge, le sexe, le code postal, la taille et la structure de la famille. Ce taux de 74 % s'avère satisfaisant en comparaison du taux de réponse rapporté par plusieurs enquêtes. Par exemple, les taux de réponse de l'enquête sur l'alimentation en Nouvelle-Écosse sont de 79,7 % parmi les répondants repérés et de 60,0 % pour l'échantillon total [23]. L'enquête Manitoba Heart Health a connu pour sa part des taux de réponse de 77,1 % parmi les répondants repérés et de 60,8 % pour l'échantillon total [24].

Évidemment, en considérant les divers types d'erreurs qui peuvent affliger un appariement de grande envergure, on comprend qu'il n'est pas réaliste de s'attendre à un taux de 100 %. Des données erronées, des décalages dans la collecte ou la mise à jour de l'information, ainsi que des différences conceptuelles entre les ensembles à appairer influencent inévitablement le taux de succès de tout appariement probabiliste. Par ailleurs, bien que les individus non appariés présentent des caractéristiques relativement différentes des individus appariés, une information socio-démographique très riche est disponible selon le recensement à propos de cette population non appariée. Ces renseignements ont servi à sélectionner un échantillon d'appariements représentatif de la population entière en vue d'effectuer des analyses de qualité.

Dans 95,5 % des cas, les paires formées associent effectivement les données d'utilisation des soins de santé d'un individu aux données socio-économiques du même individu au recensement de 1986. Ce taux d'exactitude est exceptionnel compte tenu de l'absence des noms, prénoms et jour de naissance dans le processus d'appariement.

La précision obtenue en estimant divers indicateurs liés à la consommation de soins de santé (comme la mortalité, le nombre et les coûts des services de santé, le nombre et la durée de séjour à l'hôpital) justifie le soin avec lequel les méthodes d'appariement et d'échantillonnage ont été élaborées.

A la lumière des taux d'appariement, des taux de concordance des noms et de la précision des estimations, on peut dire que la nouvelle base créée est non seulement unique au Canada, mais que, la qualité des données qui y sont codées dépasse largement celle de bien des enquêtes par interview.

Au moment où les dépenses en matière de santé dépassent les 10 % du PIB au Canada et les 13 % aux États-Unis [25], des efforts substantiels sont entrepris pour identifier les relations entre l'utilisation des soins de santé et la santé elle-même. Pendant qu'on suspecte que le niveau de santé perçu par un patient expliquerait une bonne partie de sa consommation, plusieurs études portent sur la consommation d'une clientèle particulière comme les personnes âgées ou sur la consommation des soins de santé pour les années précédant la mort [16,26,27,28].

Dans cet ordre d'idée, la base de microdonnées nouvellement construite ouvre la porte à plusieurs études qui n'ont jamais été abordées au Canada. Par exemple, un des projets défendus par le MCHPE consiste à analyser les différences de morbidité selon la profession de l'individu en examinant jusqu'à quel point l'utilisation des services de soins de santé pour une classe de maladie est reliée au groupe de base des professions. Les données du recensement permettent de classer l'individu selon la profession qu'il a déclaré, selon qu'il a un emploi ou non et selon qu'il fait partie de la population active ou non. En utilisant le volume de la classification internationale des maladies 9ième révision (CIM-9), les conditions de santé principalement étudié seront les désordres musculo-squelettique, les maladies cardio-vasculaires, les désordres mentaux, les maladies gastro-intestinales et les accidents.

D'un point de vue encore plus général, on entrevoit une analyse des différences dans le niveau d'utilisation des services de soins de santé selon le statut socio-économique mais à différents moments de la vie. D'une part, il est bien documenté que la plus grande consommation des services de santé soit vers la fin de la vie [16,27,28]. D'autre part, une diminution importante des taux de mortalité infantile entre les années 1960-1990 a aussi été observée [29a, 29b]. Ces deux événements à eux seuls justifient que des comparaisons plus poussées soient entreprises à tous les niveaux d'âge. Les données sur les visites chez les médecins, les soins de santé à domicile et les admissions à l'hôpital permettront de comparer l'utilisation en soins de santé pour les groupes d'âge selon le statut socio-économique pour les classes de maladies susmentionnées. De plus, il existe déjà des études [30,31] qui permettent de croire que les différences d'état de santé selon les statuts socio-économiques sont plus grandes chez les gens âgés entre 35 et 64 ans que pour les autres groupes d'âges. Les analyses par groupes d'âges tirées de notre nouvelle base de données pourraient confirmer ces hypothèses ou encore jeter une lumière nouvelle sur ces questions.

Même si les liens entre le statut socio-économique et la santé font l'objet de recherche intensive, l'un des problèmes les plus fréquemment rencontrés dans ce type de recherche réside dans l'impossibilité d'avoir une information précise sur le statut socio-économique au niveau des individus. Plusieurs chercheurs n'ont d'autres choix que d'utiliser un indicateur provenant de l'agrégation des données fiscales ou censitaires pour une région géographique plus ou moins grande comme le code postal ou le secteur de dénombrement [33]. Peu de recherches ont vérifié l'impact et la validité de cette méthodologie. La sensibilité de tels modèles à détecter des déterminants plus subtils mais théoriquement très importants s'en trouve d'autant réduite. Une étude des maladies respiratoires des enfants manitobains de 0 à 4 ans devrait permettre une analyse de l'impact de l'utilisation de données socio-économiques agrégées par rapport aux microdonnées. En effet, l'utilisation des services de soins de santé pour les crises respiratoires aiguës présente une différence de comportement importante selon les diverses classes de statut socio-économique. Les enfants de milieux urbains pauvres ont 4 fois plus de chance d'être hospitalisés pour des crises respiratoires aiguës que ceux des milieux les plus riches (MCHPE). Ce type de crise et les pneumonies représentent à elles seules 25 % de toutes les admissions à l'hôpital sans chirurgie pour les enfants manitobains de moins de 15 ans. De même, les taux de pontages coronariens parmi les personnes âgées présentent des caractéristiques qui permettront une analyse de l'effet de l'utilisation de données socio-économiques agrégées.

Le fichier de l'ESLA combiné aux données administratives des fichiers sur l'utilisation des soins de santé ouvre la porte à des comparaisons qui, jusqu'ici, ont été difficiles, voire impossibles à effectuer. D'un côté, l'ESLA nous fournit une image claire et détaillée des individus qui souffrent d'incapacité. Que ce soit par groupe d'âge et de sexe, par type d'incapacité (mobilité, vue, ouïe, dextérité, cognitif,...) ou par le niveau de sévérité, la population manitobaine souffrant d'une incapacité pourra être comparée à la population générale grâce au fichier du recensement. Autant, la mortalité que l'utilisation des soins de santé feront l'objet d'analyses spécifiques.

REMERCIEMENTS

Les auteurs tiennent à remercier les personnes suivantes pour leur importante et généreuse contribution à ce travail : Shelley Derksen , J. Patrick Nicol et Leonard McWilliam, Manitoba Centre for Health Policy and Evaluation.

BIBLIOGRAPHIE

- [1] WOLFSON, M.C., ROWE, G., GENTLEMAN, J.F., et TOMIAK, M. (1993). Career earnings and death: a longitudinal analysis of older canadian men. *Journal of Gerontology: Social Sciences*.
- [2] MARMOT, M.G. (1986). Social inequalities in mortality: the social environment. Dans *Class and Health, Research and Longitudinal Data*, (Éd. R G. Wilkinson). London: Tavistock Publications.
- [3] WILKINS, R., ADAMS, O., et BRANCKER, A. (1991). Changes in mortality by income in urban Canada from 1971 to 1986. *Health Reports*, 1(2), 137-174.
- [4] ANDERSON, G., GRUMBACH, K., LUTT, H., ROOS, L.L., et MUSTARD, C. (1993). Use of coronary artery bypass surgery in the United States and Canada: influence of age and income. *Journal of the American Medical Association*, 269, 1661-1666.
- [5] DOUGHERTY, G., PLESS, I.B., et WILKINS, R. (1990). Social class and the occurrence of traffic injuries and death in urban children. *Canadian Journal Of Public Health*, 81, 204-209.
- [6] GENTLEMAN, J.F., WILKINS, R., NAIR, C., et BEAULIEU, S. (1991). An analysis of frequencies of surgical procedures in Canada. *Health Reports*, 3(4), 291-309.
- [7] STATISTIQUE CANADA (1994), Health status of Canadians: Report of the 1991 General Social Survey, Cat. No. 11-612E, No. 8.
- [8] ONTARIO MINISTRY OF HEALTH (1992), Enquête Santé Ontario de 1990, highlights and working papers, Toronto.
- [9] Ministère de la Santé et des Services Sociaux (1988), Et la santé ça va ? Rapport de l'enquête Santé Québec 1987, Tome I.
- [10] STATISTIQUE CANADA (1988), The Health and Activity Limitation Survey, Selected data for Canada, provinces and territories, Cat. No. 41034.
- [11] STATISTIQUE CANADA (1984), Report of the Canadian Health and Disability Survey, Cat. No. 82-555E.
- [12] SANTÉ ET BIEN-ÊTRE SOCIAL CANADA(1981), La santé des canadiens : Rapport de l'enquête santé Canada, Cat. No. 82-538F.
- [13] STATISTIQUE CANADA (1986), *Le recensement en bref*, Cat. No. 99-104F.
- [14] STATISTIQUE CANADA (1986), *Revue générale du recensement*, Cat No. 99-137F.
- [15] ROOS, L.L., NICOL, J.P., et CAGEORGE, S.M. (1987). Using administrative data for longitudinal research: comparisons with primary data collection. *Journal of Chronical Diseases*, 40(1), 41-49.
- [16] ROOS, N.P., MONTGOMERY, P., et ROOS, L.L. (1987). Health care utilization in the years prior to death. *The Milbank Quarterly*, 65(2), 231-254.
- [17] SHAPIRO, E. et ROOS, L.L. (1984). Using health care: rural/urban differences among the Manitoba elderly. *The Gerontologist*, 24(3), 270-274.
- [18] SMITH, MARTHA E., (1981). Generalized Iterative Record Linkage System, Health Division, Statistique Canada.
- [19] FELLEGI, I.P. et SUNTER, A.B. (1969). A theory for record linkage. *Journal of the American Statistical Association*, 64, 1183-1210.
- [20] DOLSON, D., McCLEAN, K., MORIN, J.-P., et THÉBERGE, A. (1987). Plan d'échantillonnage pour l'enquête sur la santé et les limitations d'activités. *Techniques d'enquête*, 13(1), 101-117.
- [21] DAVID, P. et autres (1993). Linking Survey and Administrative Data to Study the Determinants of Health. *Proceedings of the American Statistical Association*, San Francisco.
- [22] STATISTIQUE CANADA (1994). *Causes of death 1992*, Cat. No.. 84-208.

- [23] MACLEAN, D.R. (1993). Report of the Nova Scotia Nutrition Survey. Nova Scotia Heart Health Program, Department of Health, Government of Nova Scotia.
- [24] YOUNG, T.K., GELSKEY, D.E., MACDONALD, S.M., HOOK, E., et HAMILTON, S. (1991). The Manitoba heart health survey: technical report.
- [25] SANTE ET BIEN-ETRE SOCIAL CANADA (1993), Les dépenses en matière de santé au Canada - 1991, Direction générale de la politique, de la planification et de l'information.
- [26] ROOS, N.P. (1989). How a universal health care system responds to an aging population, *J. Aging Health*
- [27] SHAPIRO, E. et ROOS, L.L. (1981). Preliminary Findings on Health Care Utilization by the Elderly Medical Care, vol. 19, No 6.
- [28] BARER M.L. et autres (1987). New evidence on old fallacies. *Soc. Sci. Med.* 24(10), 851-862.
- [29a] PAPPAS G., QUEEN S., et al. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N Engl. J. Med.* 329, 103-109.
- [29b] MARMOT M.G. and McDOWALL M.E. (1986). Mortality decline and widening social inequalities *Lancet* I, 274-276.
- [30] UGNAT A.M. et Mark E. (1987). Life expectancy by sex, age and income level. *Chronic Disease in Canada.*
- [31] WILLIAMS D.R. (1990). Socio-economic differentials in health: a review and redirection. *Social Psychology Quarterly*, 53(2):81-99.
- [32] Kovacevic M. (1995). The weight adjustment for the sample from the 'whole population database'(Private Household component), Technical note, Statistics Canada, March 24,(not published).
- [33] Geronimus A.T., Bound J., Neidert L.J. (1994). On the validity of using Census geocode characteristics to proxy socioeconomic status, Annual meetings of the Population Association of America.
- [34] Wilkins R. (1993). Use of postal codes and addresses in the analysis of health data, *Health Reports*, 5(2): 157-177, Cat. No. 82-003.

ANNEXE

Pour assurer une certaine homogénéité dans les poids qui seront utilisés lors de l'analyse, il a fallu créer une variable que l'on appelle type de dénombrement. Elle indique si les opérations du recensement ont distribué des questionnaires 2B à tous les ménages du secteur de dénombrement (SD) ou seulement à un ménage sur cinq. On y réfère par les SD «prendre tous» et les SD «prendre quelques-uns». De plus, une différence de plus de 27 % dans les taux d'appariement des SD «prendre tous» (51,8 %) par rapport aux SD «prendre quelques-uns» (78,9 %) renforçait l'importance d'utiliser cette variable pour la stratification. Notons que les SD «prendre tous» se situe surtout dans des régions éloignées ce qui explique les résultats d'appariement moins performant.

La deuxième variable construite se nomme la structure du ménage. Elle est définie à partir d'une classification hiérarchique des familles de recensement composant le ménage. Cette classification est basée sur la structure de la famille selon les catégories suivantes : ménage formé de plus d'une famille mais l'une d'elle contenant des enfants, ménage formé de plus d'une famille chacune ne contenant aucun enfant, ménage formé d'un couple avec enfant, ménage formé d'une seule famille monoparentale, ménage formé seulement d'un couple, ménage d'une seule personne jamais mariée et finalement ménage d'une seule personne ayant déjà été mariée. Le taux d'appariement variant considérablement selon la structure du ménage, cette variable s'imposait d'elle-même dans le processus de stratification.

La troisième variable est dérivée de la question 20 du questionnaire du recensement 2B. Cette question permettait aux répondants d'identifier s'il était limité dans ses activités. Si au moins un membre du ménage avait indiqué une limitation, le ménage était catégorisé «avec limitation». Autrement, on le classait «sans limitation». La volonté d'étudier les liens entre l'utilisation de soins de santé et la limitation d'activité rendait cette variable indispensable.

La quatrième variable construite est le groupe d'âge représentant le ménage. Pour conserver une définition cohérente, il fallait désigner l'individu qui représenterait le ménage. On identifiait tout d'abord la famille qui définissait la structure du ménage. S'il s'agissait d'une famille avec enfants, l'âge de la femme la plus âgée était conservé car l'âge de la mère devrait être corrélé avec le nombre d'enfant ainsi qu'avec leur âge. Dans le cas où il n'y avait pas de femmes, l'homme le plus âgé était choisi. Si la famille retenue ne contenait aucun enfant, l'adulte le plus âgé était choisi. Les limites d'âge des groupes étaient de 0 à 44 ans, de 45 à 64 ans et finalement 65 ans et plus. Une double justification nous incitait à utiliser cette variable. En effet, l'âge constitue un déterminant important de l'utilisation des soins de santé mais en plus, le taux d'appariement varie considérablement d'un groupe d'âge à un autre.

La variable sexe du ménage fut construite uniquement pour les familles monoparentales et les familles d'une seule personne. La personne qui a servi à définir le groupe d'âge désignait aussi le sexe du ménage. La comparaison des familles monoparentales et de celles possédant deux parents justifiait l'utilisation de cette variable.

La variable géographie permettait de distinguer les ménages en zone rurale par rapport à ceux en zone urbaine selon la définition du recensement.

La variable éducation est définie comme le plus haut niveau d'étude atteint par un adulte de la famille définissant la structure du ménage. Des terciles ont été calculés à l'intérieur du croisement des variables géographie et groupe d'âge et les catégories 1, 2 ou 3 ont été dérivées.

Finally, the variable mesure de revenu équivalent (low income measure) is constructed from the total household income divided by the number of individuals adjusted for consumption. The first individual has a value of 1, the second of 0,37, the third of 0,36, the fourth of 0,26 and finally the fifth and each of the others have a value of 0,18. As for the previous variable, these are the terciles according to the same crossings that have been preserved.