

Couplage des données du Recensement de la population de 2006 et de la Base canadienne de données sur la mortalité

Mohan B. Kumar, Rose Evra¹

Résumé

Les ensembles de données de la statistique de l'état civil, comme la Base canadienne de données sur la mortalité, n'ont pas d'identificateurs pour certaines populations d'intérêt, comme les Premières Nations, les Métis et les Inuits. Le couplage d'enregistrements entre les données de la statistique de l'état civil et les ensembles de données d'enquête ou autres ensembles de données administratives peuvent permettre de contourner cette limite. Le présent document décrit un couplage de la Base canadienne de données sur la mortalité et du Recensement de la population de 2006, ainsi que l'analyse prévue à partir des données couplées.

Mots-clés : couplage d'enregistrements; déterministe hiérarchique; statistique de l'état civil; taux de mortalité.

1. Introduction

1.1 Description

La Base canadienne de données sur la mortalité (BCDM) est un ensemble de données de la statistique de l'état civil qui comprend des renseignements sur les décès qui se sont produits au Canada (Statistique Canada, 2015). Toutefois, elle ne comporte pas d'identificateurs pour certaines populations d'intérêt, comme les Premières Nations, les Métis et les Inuits. Cela peut être contourné en couplant des enregistrements de la BCDM aux enregistrements d'enquête ou d'autres dossiers administratifs. Pour identifier les Premières Nations, les Métis et les Inuits, une des options consiste à effectuer le couplage avec le Recensement de la population (recensement). Cela a été fait précédemment avec le Recensement de 1991 (Wilkins, 2001); toutefois, comme les données du recensement ne comprennent pas de noms, elles ont d'abord dû être couplées aux données fiscales pour obtenir des renseignements nominaux. Dans le Recensement de 2006, pour la première fois, les noms ont été conservés, ce qui a permis un couplage direct entre celui-ci et la BCDM. Ce couplage permettra non seulement d'identifier les personnes qui se désignent elles-mêmes comme membres des Premières Nations, Métis et Inuits dans le questionnaire détaillé du Recensement de 2006, mais il donnera accès aux données sociodémographiques du recensement à des fins d'analyse.

Ce projet vise à coupler les données du Recensement de 2006 et celles de la BCDM, afin de permettre l'identification des membres des Premières Nations, des Métis et des Inuits décédés. Le couplage sera amélioré en tirant parti des données fiscales d'un autre ensemble de données couplées, la Base de données combinées sur la mortalité. Les données de couplage serviront à estimer et à comparer les taux de mortalité pour les populations mentionnées précédemment et les populations non autochtones au Canada.

Le couplage d'enregistrements a été entrepris par la Division de la statistique sociale et autochtone de Statistique Canada, en 2015. L'approbation officielle du couplage d'enregistrements a été accordée en 2015-2016, et le couplage a été effectué au premier trimestre de 2016-2017. Il a été mené avec l'aide des méthodologistes de la Division des méthodes d'enquêtes auprès des ménages (DMEM). L'évaluation de la qualité et la validation du couplage d'enregistrements ont été entreprises au premier trimestre de 2016-2017.

¹Mohan B. Kumar, (mohan.kumar@canada.ca); Rose Evra, (rose.evra@canada.ca); Division de la statistique sociale et autochtone, Statistique Canada, 100 promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6

2. Méthodes

2.1 Ensembles de données

Recensement de la population de 2006 : Le Recensement de la population au Canada est effectué tous les cinq ans. Il permet de recueillir des données sur les caractéristiques démographiques, sociales et économiques des Canadiens. Il comprend un questionnaire abrégé et un questionnaire détaillé. En 2006, ce dernier a été administré à un ménage sur trois, les autres ayant reçu le questionnaire abrégé. L'ensemble de données du Recensement de 2006 comporte 30,1 millions d'enregistrements, y compris certains enregistrements en double. Il y a sous-dénombrement dans le recensement, particulièrement en ce qui a trait aux jeunes hommes (Statistique Canada, 2010).

Base canadienne de données sur la mortalité (BCDM) : La BCDM est une base de données administratives qui comprend des données médicales (cause du décès) et démographiques recueillies annuellement auprès de tous les registres provinciaux et territoriaux de la statistique de l'état civil sur tous les décès au Canada. Pour la période de l'étude, 1,4 million d'enregistrements de personnes décédées entre 2006 (année du Recensement de 2006) et 2011 (dernière année de données disponible dans l'ensemble de données) sont dans le champ de l'enquête. La BCDM comporte un certain sous-dénombrement et surdénombrement (Statistique Canada, 2015) : dans le premier cas, cela vient du fait que seulement les décès qui se sont produits au Canada sont enregistrés par la statistique de l'état civil, par conséquent, les décès de Canadiens, y compris les militaires ou ceux vivant à l'étranger, ne sont pas enregistrés, et en raison d'enregistrements tardifs. En ce qui a trait au surdénombrement, il peut être le résultat de l'inclusion de non-Canadiens qui sont arrivés après le jour du recensement et qui sont décédés au Canada.

2.2 Méthodes de couplage

Les enregistrements des deux bases de données ont été couplés au moyen d'une méthode déterministe hiérarchique (Rotermann, 2014). Cela comprend l'appariement d'enregistrements entre les deux ensembles de données au moyen de variables communes. Après une évaluation initiale de leur qualité, en vue de leur utilisation pour le couplage d'enregistrements, les variables suivantes ont été utilisées : nom, sexe, code postal et date de naissance. La qualité et la disponibilité de ces variables différaient dans les deux ensembles de données. Par exemple, alors que les répartitions selon le sexe étaient similaires dans les deux ensembles, les données sur le code postal étaient manquantes dans une petite partie des enregistrements de la BCDM (0,9 %), cette proportion augmentant la dernière année où les enregistrements de la BCDM étaient disponibles (2,1 %).

Le couplage a été effectué en vagues, les critères de couplage devenant de moins en moins rigides à chaque vague subséquente (tableau 3.1-1). Au total, 20 vagues de couplage d'enregistrements ont eu lieu. Les vagues 1 à 16 ont utilisé des combinaisons et des niveaux de rigidité différents pour le couplage des variables. Au cours des quatre dernières vagues, on a utilisé les données fiscales pour le couplage. Par exemple, à la vague 1, les enregistrements devaient correspondre exactement pour les quatre variables de couplage; à la vague 4, le critère pour la variable du nom a été relâché au moyen du logiciel MixMatch; et, à la vague 9, les critères pour les variables du nom et du code postal ont été relâchés (en utilisant MixMatch pour les noms et en utilisant seulement le premier caractère du code postal). Seuls les enregistrements non couplés sont utilisés dans les vagues subséquentes.

3. Résultats préliminaires

3.1 Taux de couplage

Parmi les 1,4 million d'enregistrements de la BCDM compris dans le champ, 89 % ont été couplés aux enregistrements du recensement. Le taux de couplage², défini par le pourcentage d'enregistrements de la BCDM couplés, diffèrait selon la vague. Les vagues 1, 2, 4 et 9 ont contribué à 81 % de tous les couplages (tableau 3.1-1).

Tableau 3.1-1 : Pourcentage du total des enregistrements couplés et pourcentage des enregistrements de la BCDM couplés selon la vague

Vagues	Variables	% de couplage	BCDM
1	Nom, DN, SEXE, CP	26,28 %	23,29 %
2	Nom, DN, SEXE	12,87 %	11,41 %
3	Nom, DN, CP	0,15 %	0,13 %
4	Nom de famille, prénom (MM*) DN, SEXE, CP	27,42 %	24,30 %
5	Nom + DN + CP3 ou CP1	0,05 %	0,04 %
6	Nom + DN6 + SEXE ± CP3	1,84 %	1,63 %
7	Nom + AN + SEXE + CP	0,62 %	0,55 %
8	Nom(MM) + DN + SEXE + CP	7,71 %	6,83 %
9	Nom(MM) + DN + SEXE + CP1	14,30 %	12,67 %
10	Nom(MM) + DN + CP3	0,24 %	0,22 %
11	Nom + SEXE + CP	1,06 %	0,94 %
12	Nom(MM) + DN (au moins 5 chiffres communs) + CP + SEXE	3,26 %	2,89 %
13	Nom(MM) + AN/MN + SEXE + CP3	0,31 %	0,27 %
14	Nom et prénom (fname) + DN(une différence d'un chiffre permise) + SEXE	0,59 %	0,52 %
15	Nom(MM) + DN + SEXE	1,45 %	1,29 %
16	Nom de famille du père ou de la mère inclus dans le nom + DN + CP	0,13 %	0,11 %
17	Nom des dossiers fiscaux + DN des dossiers fiscaux + CP des dossiers fiscaux + SEXE	1,09 %	0,96 %
18	Nom des dossiers fiscaux + DN des dossiers fiscaux + SEXE	0,24 %	0,21 %
19	Nom des dossiers fiscaux (MM) + DN des dossiers fiscaux + CP des dossiers fiscaux + SEXE	0,17 %	0,15 %
20	Nom des dossiers fiscaux (MM) + CP3 des dossiers fiscaux + DN des dossiers fiscaux	0,21 %	0,19 %
Total		100,00 %	88,60 %

Notes : DN : Date de naissance, AN : année de naissance; MN : mois de naissance, DN des dossiers fiscaux : date de naissance des dossiers fiscaux, CP : code postal, CP1 : premier caractère seulement du code postal; CP3 : trois premiers caractères seulement du code postal; fname : prénom, * MM : couplages dans lesquels la restriction concernant la variable du nom a été relâchée au moyen de la version 1.3 du logiciel MixMatch, ±CP3 dans la vague 6 désigne le couplage avec et sans les trois premiers caractères du code postal.

² Le taux de couplage dans ce cas est défini comme la proportion d'enregistrements de la BCDM qui ont été couplés aux enregistrements du Recensement.

4. Évaluation de la qualité du couplage des enregistrements

4.1 Validation interne

Pour la validation interne du couplage des enregistrements, les couplages en double, les couplages avec des enregistrements hors du champ et les tendances des taux de couplage ont été examinés.

Qualité des couplages : dans le cas des enregistrements comportant une date de naissance valide du recensement, moins de dix avaient une date de décès antérieure à la date de naissance. Le sexe des personnes différait dans les deux ensembles de données pour 0,5 % des couplages.

En 2006, parmi tous les décès qui se sont produits avant le jour du recensement, 5 % ont été couplés aux dossiers du recensement. Parmi eux, 40 % se sont produits au mois de mai; il pourrait s'agir de couplages légitimes, étant donné qu'un dénombrement anticipé a eu lieu dans les réserves et dans le Nord. Parmi les décès qui avaient trait à des personnes dont la date de naissance était postérieure au jour du recensement, 1 % ont été couplés aux dossiers du recensement. Il pourrait s'agir de faux couplages; toutefois, un dénombrement tardif a eu lieu, ce qui laisse supposer que certains de ces couplages sont peut-être des couplages véritables.

Couplages en double : certains enregistrements de la BCDM ont été couplés à plus d'un dossier du recensement (à l'intérieur d'une vague). Presque tous semblaient être des doubles dans l'ensemble de données du recensement. En outre, certains dossiers du recensement étaient couplés à plus d'un enregistrement de la BCDM, certains peut-être en raison de doubles dans la BCDM, et d'autres, en raison d'une erreur de couplage.

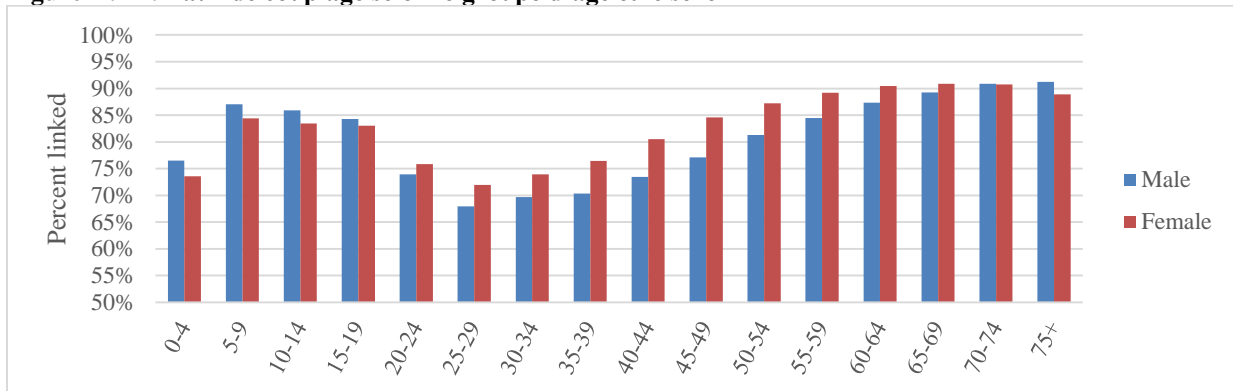
Tendances des taux de couplage : les taux de couplage ont été examinés selon des catégories de variables différentes, comme le groupe d'âge, la cause du décès, les quintiles de revenu du quartier, etc., afin de déterminer si les tendances attendues se concrétisaient.

Lorsque les taux de couplage ont été examinés selon le sexe, contrairement aux attentes, parce que les femmes sont plus susceptibles de changer de nom de famille au moment du mariage, les taux de couplage étaient plus élevés chez les hommes. Toutefois, il convient de souligner que l'utilisation des noms des dossiers fiscaux a permis le recours à des versions multiples des noms des personnes et des noms des parents; cela peut avoir amélioré les taux de couplage chez les femmes. Par ailleurs, le sous-dénombrement des hommes dans le recensement peut avoir aussi contribué aux faibles taux de couplage pour ces derniers.

Lorsque les taux de couplage ont été examinés selon l'année de décès, ils ont diminué légèrement au fil du temps, pour passer de 89 % en 2007 à 88 % en 2011, comme il fallait s'y attendre en raison de la mobilité.

Les taux de couplage différaient selon le groupe d'âge, les personnes de 25 à 44 ans affichant les taux les plus faibles (77 %) comparativement aux autres groupes. Les personnes âgées de 65 à 84 ans avaient le taux de couplage le plus élevé (90 %). Les faibles taux de couplage dans le premier groupe d'âge peuvent être expliqués par leurs taux de couverture plus faible dans le recensement, en raison peut-être d'une mobilité plus grande. Dans le groupe des 25 à 44 ans, parmi les hommes et les femmes, les taux de couplage étaient les plus faibles chez les 25 à 29 ans (figure 4.1-1).

Figure 4.1-1. Taux de couplage selon le groupe d'âge et le sexe

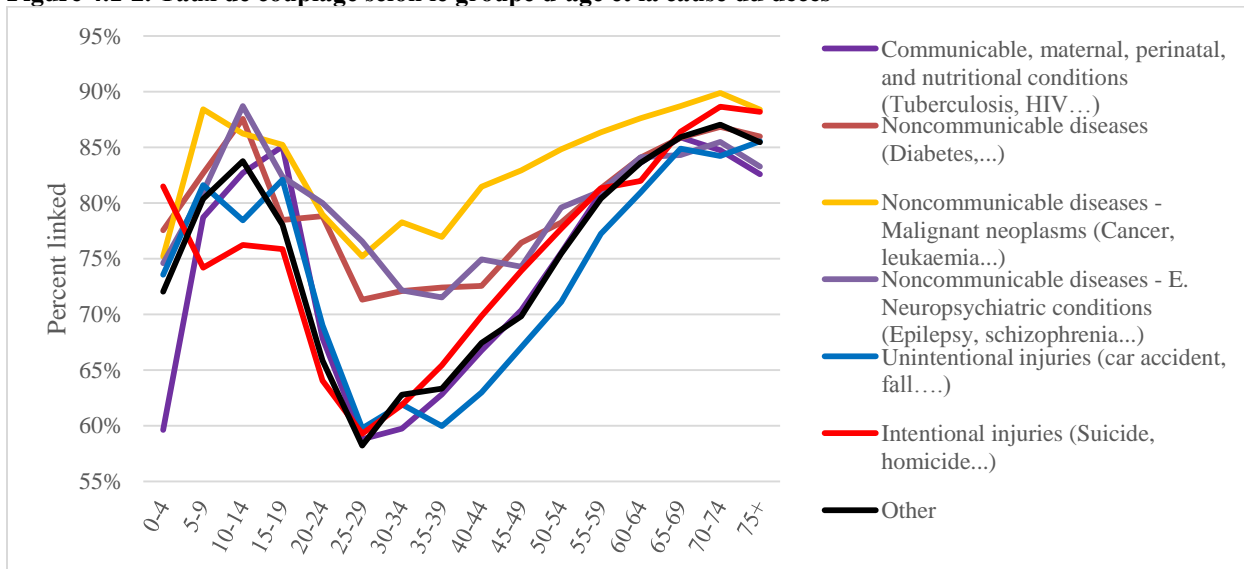


Les taux de couplage étaient plus faibles chez les personnes qui vivaient dans des quartiers à faible revenu, selon les données du code postal de la BCDM, que chez celles des quartiers à revenu élevé.

En outre, comme prévu, les taux de couplage variaient selon l'état matrimonial, les personnes mariées affichant les taux de couplage les plus élevés (93 %) et les personnes seules, les plus faibles (82 %).

Les taux de couplage ont aussi été examinés selon la cause de décès : ceux qui sont décédés par suite de blessures intentionnelles étaient les moins susceptibles d'être couplés (78 %), et ceux qui sont décédés par suite de cancers étaient plus susceptibles d'être couplés (91 %; figure 4.1-2). Les variations du couplage selon le groupe d'âge, qui sont décrites précédemment, lorsqu'elles sont examinées selon la cause du décès, ont révélé que les 20 à 35 ans qui sont décédés par suite de blessures intentionnelles ou non intentionnelles avaient les résultats de couplage les plus faibles. Comme il fallait s'y attendre, ce sont les causes de décès parmi les plus probables dans ce groupe d'âge.

Figure 4.1-2. Taux de couplage selon le groupe d'âge et la cause du décès



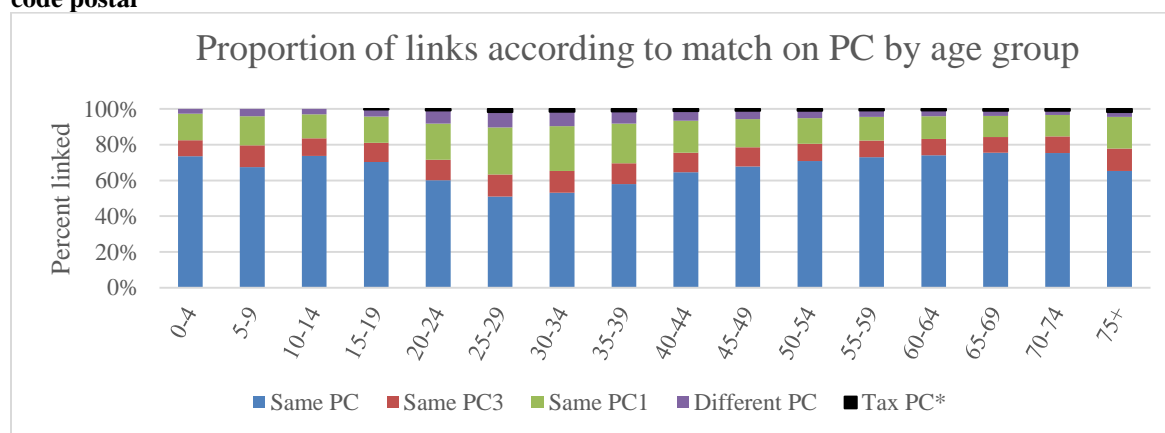
Parmi les provinces et les territoires, ces derniers affichaient les taux de couplage les plus faibles, soit près de 80 %. Le taux de couplage était le plus élevé au Nouveau-Brunswick, à 91 %. Des taux de couplage plus faibles se retrouvaient dans les provinces comptant une proportion plus forte de jeunes adultes.

Les personnes qui sont nées à l'extérieur du Canada étaient moins susceptibles d'être couplées, des taux de couplage particulièrement faibles étant obtenus chez les moins de 35 ans (61 % ou moins). Certaines de ces personnes peuvent

avoir immigré au Canada après le jour du recensement, ou peuvent avoir été des visiteurs ou des travailleurs temporaires, ce qui expliquerait le faible couplage.

Une part importante des taux de couplage plus faibles pour les 20 à 39 ans peut possiblement être expliquée par la mobilité plus grande des personnes de ce groupe d'âge. Afin d'explorer cela davantage, les taux de couplage ont été examinés selon les vagues, dans lesquelles la rigueur des critères d'appariement pour le code postal variait. Ces vagues comprenaient celles où les enregistrements devaient correspondre exactement pour l'ensemble du code postal (« même CP » dans la figure 3), correspondre exactement pour les trois premières caractères (« même CP3 »), correspondre exactement pour le premier caractère seulement (« même CP1 »), et celles où le code postal pouvait différer dans les deux enregistrements (figure 4.1-3). Dans le cas des groupes des 25 à 29 ans, des 30 à 34 ans et des 35 à 39 ans, un pourcentage plus faible a été couplé dans ces vagues, qui devaient correspondre entièrement pour l'ensemble du code postal, ce qui laisse supposer que la mobilité a joué un rôle dans les faibles taux de couplage pour ces groupes d'âge.

Figure 4.1-3. Taux de couplage selon les vagues comportant différents niveaux de rigueur pour la variable du code postal



D'autres exercices de validation et examens administratifs sont en cours, et les résultats serviront à évaluer la pertinence de l'utilisation de l'ensemble de données à des fins d'analyse.

5. Analyses des données couplées

5.1 Objectifs

Le couplage d'environ 90 % des enregistrements de la BCDM avec les dossiers du recensement permet l'identification des membres autodéclarés des Premières Nations, des Métis et des Inuits dans les enregistrements de la mortalité, ce qui permet l'estimation des taux de mortalité dans ces populations et dans la population non autochtone. Il offre aussi la possibilité de tenir plusieurs autres analyses dans ces populations et d'autres populations d'intérêt.

Les objectifs particuliers de l'analyse sont les suivants : 1) produire des estimations des taux de mortalité fondées sur la population pour les Premières Nations, y compris les Indiens inscrits, les Inuits et les Métis, ainsi que les non-Autochtones, toutes les causes combinées et pour des causes particulières, y compris la maladie cardiovasculaire, le cancer, les maladies respiratoires et le suicide, 2) estimer les espérances de vie à la naissance, à l'âge de 21 ans et à l'âge de 65 ans pour les populations mentionnées précédemment, et 3) évaluer les répercussions des caractéristiques socioéconomiques et démographiques sur le risque de décès, pour toutes les causes et des causes différentes, pour les Premières Nations, les Métis et les Inuits comparativement aux non-Autochtones.

Parmi les autres projets possibles figurent les suivants : 1) estimation des espérances de vie des Premières Nations, des Métis, des Inuits et des populations non autochtones, 2) examen de la mortalité évitable parmi ceux qui sont décédés prématurément (avant l'âge de 75 ans), 3) estimation des années possibles de vie perdues pour ceux qui sont

décédés prématurément, 4) validation des résultats de la mortalité parmi les Indiens inscrits, à partir des données du Registre des Indiens, 5) examen du rôle des facteurs socioéconomiques et démographiques sur les taux de décès ou les disparités entre les taux de mortalité de nombreuses populations d'intérêt et de la population non autochtone, 6) élaboration d'analyses similaires pour les immigrants et les minorités visibles dans le cadre de rapports, de feuillets d'information et de tableaux de données, et 7) prolongement de ces analyses, à partir d'années additionnelles de données de la BCDM au fil du temps.

Bibliographie

Rotermann M, Sanmartin C, Carrière G, Trudeau R, St-Jean H, Saïdi A, et al (2014), "Two approaches to linking census and hospital data", Health Reports, 25(10), pp. 3–14.

Statistics Canada (2010), "Table 1.2.2 Estimated population net undercoverage and standard errors for various characteristics, 2006 Census", available from: http://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/coverage-couverture/tbl/tbl-01_2_2-eng.cfm

Statistics Canada (2015), "Vital Statistics – Death Database", available from: <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3233>

Wilkins R, Tjepkema M, Mustard C, Choinière R. (2008), "The Canadian census mortality follow-up study, 1991 through 2001", Health Reports, 19(3), pp. 25–43.