



Statistique
Canada

Statistics
Canada

L'appariement dans les études de santé à Statistique Canada

www.statcan.gc.ca



L'histoire du Canada
racontée en chiffres

Philippe Finès, Ph. D
Statistique Canada

CIQSS

3535, chemin Queen-Mary, local 420

Université de Montréal, 13 avril 2017

Canada



Sommaire

- Définition et Principes
- Défis et mesures de qualité
- Processus détaillé
- L'appariement ESCC – Mortalité – Congés des patients
 - Bases de données
 - Étapes
 - Résultats
- Autres appariements
- Articles et résultats
- Adresses utiles
- Conclusions



Définition et Principes

Qu'est-ce que l'appariement?

- C'est la mise en relation d'enregistrements de fichiers différents et qui concernent la même entité, par exemple la même personne
- On effectue l'appariement en utilisant une information unique commune aux fichiers
 - Ex: nom, date de naissance, code postal, identificateurs personnels (numéro d'assurance santé ou d'assurance sociale)
- Les méthodes d'appariement sont avant tout des méthodes “exactes”: elles associent la même unité (et non une unité similaire)
 - Déterministes
 - Déterministes hiérarchiques
 - Probabilistes



Pourquoi appairier?

- Chaque base de données utilisée dans l'appariement est incomplète
 - Données d'enquêtes: beaucoup de données socioéconomiques, mais pas d'issues de santé (mortalité, morbidité);
 - Données administratives: issues de santé (hospitalisation, morbidité) mais information individuelle limitée (ex: pas de statut autochtone, pas d'immigration, pas d'éducation, etc...)
- Appairier les données permet d'obtenir un fichier complet qui comprend à la fois les données individuelles et les issues de santé.
 - On peut établir des liens entre l'issue (ex: hospitalisation) et les données individuelles (ex: habitudes de vie)
- Études populationnelles sur un échantillon représentatif de Canadiens
- Grands échantillons – permettent d'étudier des populations spécifiques et des événements rares
- Possibilité de faire des comparaisons entre juridictions

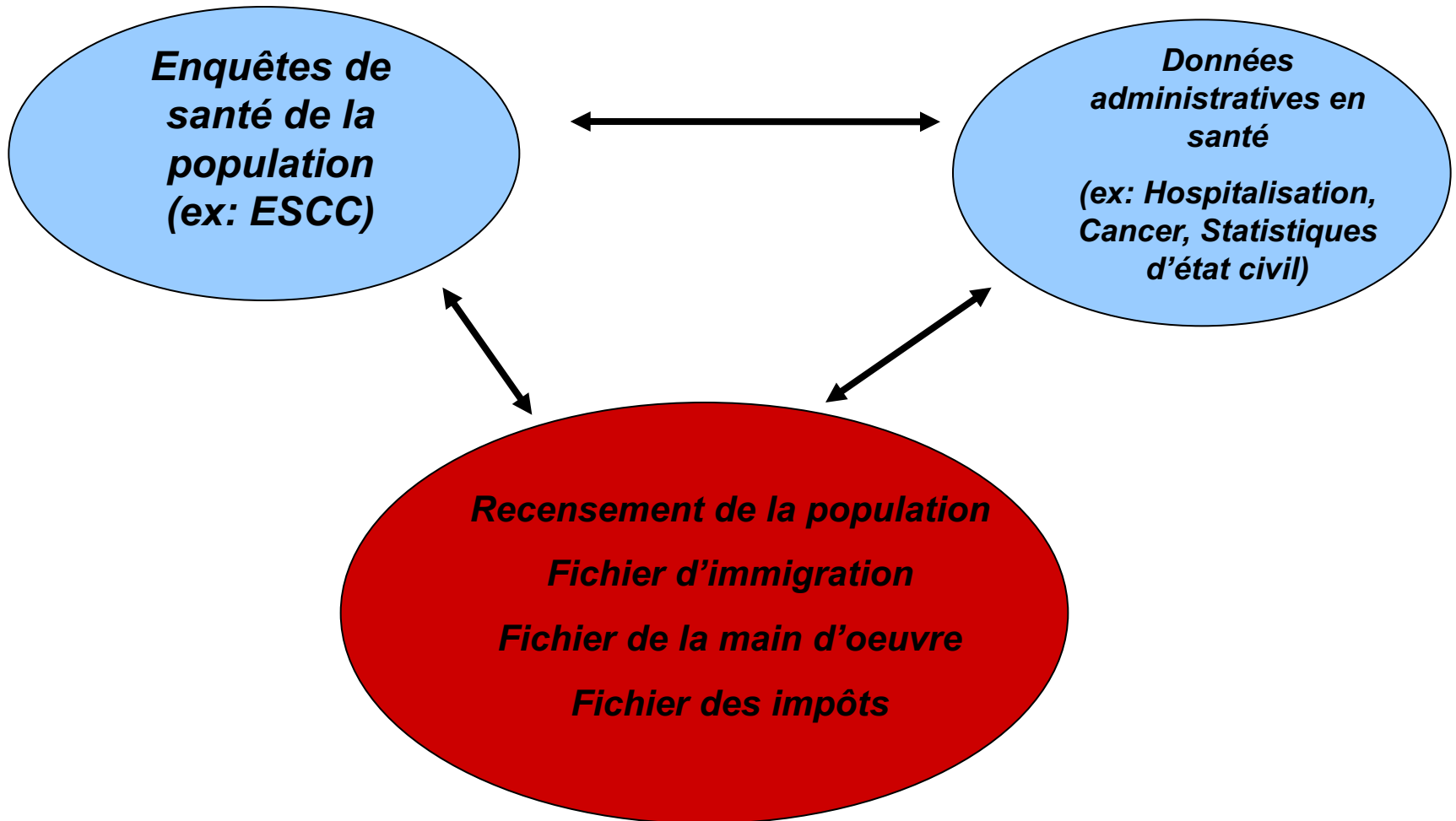


Exemples de projets de recherche

1. Comprendre l'interaction entre les facteurs de risque socioéconomiques et les comportements, et leur effet sur l'utilisation et les coûts de services hospitaliers
2. Déterminer comment les différences dans la prévalence des facteurs de risque expliquent la variation dans l'utilisation des services hospitaliers
3. Examiner l'interaction entre les facteurs de risque, l'exposition à la pollution de l'air, la mortalité et l'utilisation des services hospitaliers



Appariements en santé: Fichiers de données





Appariement des données de santé – Pourquoi à Statistique Canada?

- Statistique Canada est à l'avant-garde dans le développement de méthodologies et logiciels pour l'appariement (G-Link)
- *La loi sur la statistique* assure une protection de l'information confidentielle
- Analyse de l'impact de l'appariement sur la privauté
 - Les bénéfices pour l'analyse statistique doivent dépasser l'invasion de la privauté
 - Il faut trouver un bon équilibre
- Possibilité de rendre les données appariées accessibles dans les *Centres de Données de Recherche*
 - *Plus de détails à la fin de la présentation*



Forces, Limites, Défis



Forces

- Utilisation de sources de données existantes – valeur ajoutée avec une approche efficace
- Grandes cohortes nationales – permettent l'étude de maladies rares, l'analyse de sous-groupes, les comparaisons entre juridictions
- Le groupe de comparaison est construit durant le même appariement
 - On construit en même temps le groupe des cas (=individus avec l'issue à l'étude) et le groupe contrôle
- Possibilité d'analyser les déterminants sociaux de la santé en conjonction avec d'autres variables – telles que le statut socioéconomique, les facteurs de risque individuels ou écologiques, etc.



Limites

- Exclusion de certaines populations:
 - Recensement: Réserves incomplètement énumérées; Population en institution exclue
 - Enquête sur la Santé dans les Collectivités Canadiennes – Population autochtone exclue
- Les taux d'appariement peuvent varier d'un groupe à l'autre – Études de validation nécessaires
- Coût lié à l'appariement
- Expertise requise pour effectuer l'appariement



Défis

- 1. Variables utilisées pour l'appariement
 - Que faire si les variables communes aux fichiers sont insuffisantes?
 - Variables utilisées en général pour l'appariement:
 - Ex: nom, date de naissance, code postal, identificateurs personnels (numéro d'assurance santé ou d'assurance sociale)
 - Mais ces variables ne sont pas toujours présentes dans tous les fichiers
- 2. Résultats concernant les taux d'appariements – Validation interne
 - Qu'est-ce qu'un “bon” taux d'appariement?
 - On vise 100% lorsque la même population est présente dans tous les fichiers
 - Issues de santé – hospitalisation, cancer non présents chez tout le monde (la mort, par contre, oui)
- 3. Validation externe
 - Les données sont-elles “bonnes” pour la recherche?
 - Analyse de couverture pour déterminer si les données appariées présentent le taux “espéré” d'issues de santé dans la population d'intérêt



Validation

- Deux types de validation:
 - Validation interne – qualité de l'appariement (**taux d'erreur**)
 - *Est-ce que les enregistrements appariés sont corrects?*
 - *Parmi les enregistrements non appariés, a-t-on oublié des variables d'appariement?*
 - Validation externe – qualité des données appariées (**représentativité** du fichier)
 - *Est-ce que les issues observées dans le fichier apparié représentent la population d'intérêt?*

Validation interne

- Comparer les résultats de la révision manuelle à ceux de la solution obtenue dans l'appariement

Révision manuelle	Appariement	
	Liens	Non-liens
	Liens a (Vrais positifs)	b (Faux négatifs)
Non-liens	c (Faux positifs)	d (Vrais négatifs)

- Calculer les taux:
 - Faux positifs (on croit que c'est un lien; mais ce n'en est pas un) → c
 - Faux négatifs (on croit que ce n'est pas un lien; mais c'en est un) → b
 - Sensibilité = Vrais positifs / Total des positifs = $a / (a+c)$
 - Spécificité = Vrais négatifs / Total des négatifs = $d / (b+d)$



Validation interne

- Critère de Décision
 - Appariements “rigoureux”, “précis” au risque de perdre des enregistrements (stratégie d’appariement conservatrice)
 - Objectif: diminuer a , diminuer c → augmenter spécificité, diminuer sensibilité
 - Appariements lâches, au risque d’avoir des appariements inadéquats
 - Objectif: augmenter b , augmenter d → augmenter sensibilité, diminuer spécificité
 - Que faire p.ex si le nom et le sexe sont les mêmes mais pas le code postal?



Validation externe

- On crée le fichier apparié ou *analytique*
- “Utilisabilité” en analyse et en recherche
 - *Les données vont-elles produire des résultats non biaisés et généralisables?*
- Il n’y a pas beaucoup de normes; plusieurs méthodes sont utilisées à Statistique Canada.
- Mais en gros les méthodes procèdent toutes comme suit ...



Validation externe

- Validité intuitive – Expertise sur le sujet
 - Produire les résultats sur les taux (incidence ou prévalence) de l'issue (hospitalisation, décès, etc) analysée
 - Comparer les taux du fichier apparié à ceux disponibles dans la population (revue de littérature)
 - Effectuer l'analyse de couverture – proportion dans le fichier apparié des événements observés dans la population
 - Ex: le fichier apparié comprend 10% des hospitalisations pour asthme. Malgré cela, on s'attend (en gros) à ce que le taux d'hospitalisation pour asthme dans le fichier apparié soit comparable à celui observé dans la population
 - Examiner les premiers résultats de recherche
- Un travail de recherche est en cours à Statistique Canada pour développer une méthodologie de validation des appariements



Processus détaillé



Le processus pour appariement des données

Phase 1: Planification du projet

- Spécifier les besoins
- Concevoir une stratégie d'appariement de données
- Faire approuver le projet d'appariement des données

Phase 2: Appariement d'enregist- rements

- Préparation des données – variables d'appariement
- Enregistrement d'appariement
- Évaluer la qualité – validation interne et externe

Phase 3: Intégration, Analyse, Accès

- Créer et document le fichier apparié;
- Effectuer l'analyse
- Assurer l'accès (en particulier dans les CDR) et l'archivage
- Évaluer le projet d'appariement



Approbation du projet d'appariement

- Les appariements doivent satisfaire un protocole dont le point d'orgue est la révision par le comité des politiques – présidé par le Statisticien en chef
- Un projet d'appariement doit satisfaire les critères suivants:
 - Importants bénéfiques pour les Canadiens de l'information obtenue avec le projet
 - Aucun autre moyen de répondre à la question de recherche
 - Confidentialité des données individuelles assurée
 - Normes rigoureuses pour la méthodologie et la qualité des données
- Les projets d'appariements sont annoncés publiquement dans le site de Statistique Canada
- Processus continu développé pour permettre une nouvelle utilisation de données appariées existantes

L'appariement ESCC – Mortalité – Congés des patients



Enquête sur la santé dans les collectivités canadiennes (ESCC) appariée à la base de données sur la mortalité et à la Base de données sur les congés des patients

Financement:
Santé Canada, Université McGill, ICES/OHRI

Statistique Canada • Statistics Canada

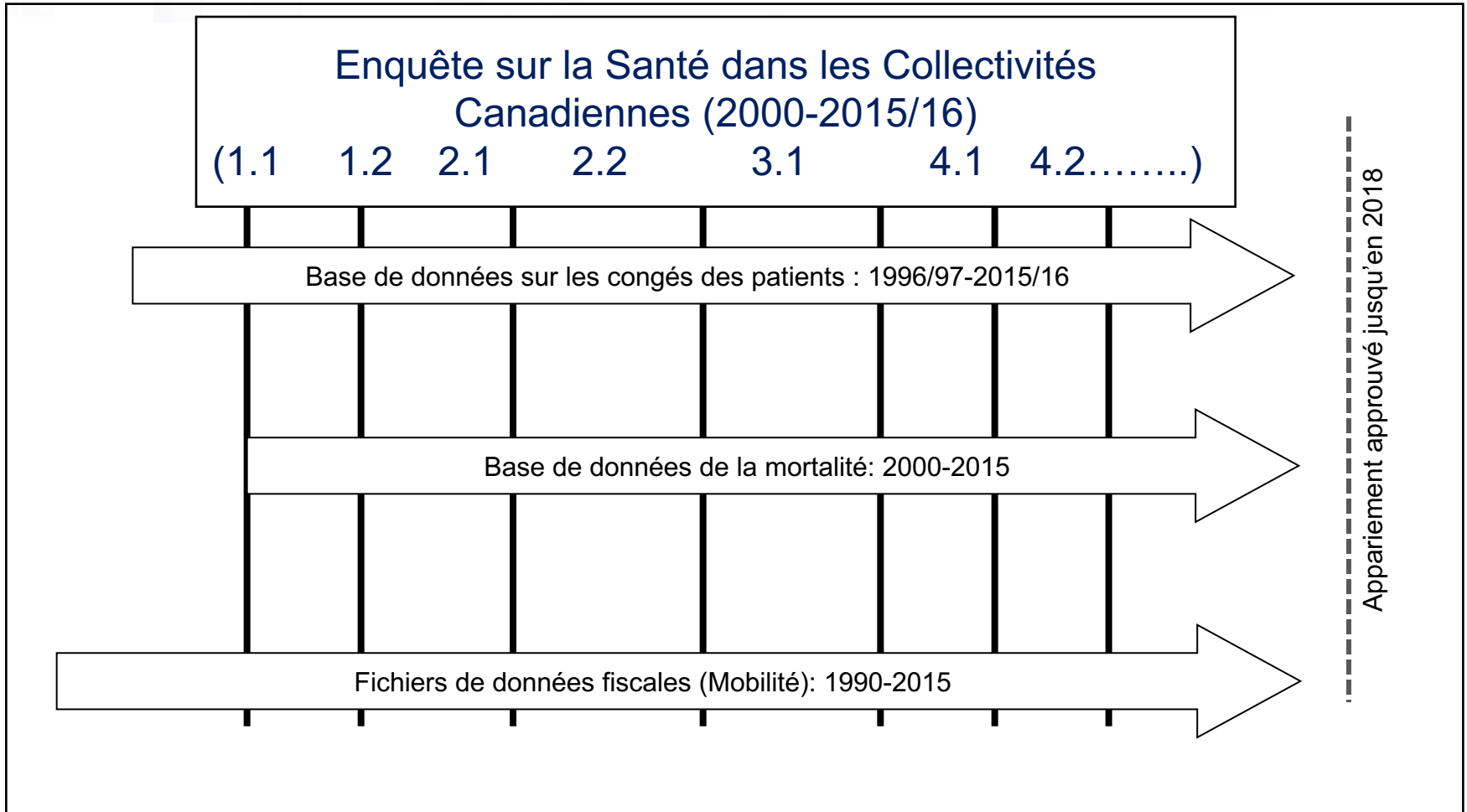


Contexte – Pourquoi un appariement?

- Intérêt grandissant pour évaluer l'impact des habitudes de vie – tabagisme, nutrition, activité physique – sur les issues de santé et l'utilisation des services de santé
 - Besoin d'utiliser une approche multivariée étant donné les nombreux déterminants sociaux de la santé
- Intérêt pour la mesure de l'impact des données environnementales (i.e. pollution de l'air) sur les issues de santé en contrôlant les facteurs de risques concurrents (i.e. tabagisme)



Appariement ESCC (approuvé)





Fichiers

- ESCC: Enquête sur la santé dans les collectivités canadiennes (2000-2011)
 - .1 et annuelles; .2: Santé mentale, Nutrition, Vieillessement en santé
 - Fichier des appariements à partager: Répondants ESCC qui ont accepté le partage et l'appariement
 - Enquête transversale biennale (~130,000); annuelle (~65,000) après 2007;
 - Représente les personnes de 12 ans et plus, couvre environ ~98%
 - Exclut membres des forces armées, personnes en institution, personnes dans les réserves, personnes dans certaines régions éloignées
 - Contenu principal: État de santé, facteurs de risque, conditions chroniques, indicateurs socioéconomiques
 - Contenu par thème depuis 2002 (~30,000)
 - Santé mentale (Cycle 1.2), nutrition (Cycle 2.2), vieillissement(Cycle 4.2)



Fichiers

- Base de données sur les congés des patients (1996-2013/14)
 - Données de congé surtout pour les hôpitaux de soins de courte durée (~3 millions d'enregistrements/an)
 - Provient de l'ICIS (Institut Canadien d'information sur la santé)
 - Contient des données démographiques, non médicales et cliniques
 - Contient de l'information sur les coûts
 - Information sur les événements et sur les personnes
 - Le Québec est exclu de ce fichier
 - Appariement probabiliste(G-Link)
 - Date de naissance, sexe, code postal, province



Fichiers

- Base de données sur la mortalité canadienne (2000-2011)
 - Décès parmi les personnes de 12 ans et plus (n=2.8 millions)
 - Contient la cause de décès
 - Appariement probabiliste(G-Link)
 - Noms, date de naissance, code postal

- Fichiers de données fiscales (1996-2011)
 - Formulaires de déclaration d'impôt pour tous les contribuables (n=33.5 millions)
 - Lieu annuel de résidence (code postal) *–pas de données de revenu!*
 - Appariement probabiliste(G-Link)
 - Noms, date de naissance, code postal



Étape 1: Préparation des données

- But: préparer les données des fichiers sources pour l'appariement
- Eligibilité des observations et des répondants pour l'appariement
 - Identifier les observations éligibles pour l'appariement;
 - Comparer les caractéristiques des éligibles et des non-éligibles
 - Évaluer le biais
 - Calculer les nouveaux poids (basés seulement sur les individus appariés)



Étape 1: Préparation des données

- Qualité et complétude des variables d'appariement
 - Variables: noms, date de naissance, code postal, sexe, identificateurs uniques (Assurance-santé, Assurance sociale)
 - Taux d'entrées manquantes et invalides
- Pouvoir discriminatoire des variables d'appariement:
 - Valeur unique des données pour justifier la stratégie d'appariement
 - Mesure d'entropie (Shannon)
- On doit calculer le taux d'éligibilité par cycle, par variable

ESCC – Éligibilité pour appariement

Répondants ESCC éligibles pour l'appariement, par année d'enquête, 2000/02 à 2011			
Année, cycle d'ESCC	Nombre de répondants	Répondants qui ont accepté le partage et l'appariement	
		n	%*
Cohorte 1			
2000/02 (1.1)	131,535	117,837	89.6
2003/04 (2.1)	135,573	112,851	83.2
2005/06 (3.1)	132,947	113,880	85.7
2007	65,946	57,083	86.6
2008	66,013	55,592	84.2
2009	61,679	52,475	85.1
2010	63,191	52,198	82.6
2011	63,542	52,858	83.2
Total	720,426	614,774	85.3
Cohorte 2			
2002 (1.2 - Santé mentale)	36,984	31,720	85.8
2003 (2.2 - Nutrition)	35,107	29,897	85.2
2009 (4.2 - Vieillesse en santé)	30,865	25,486	82.6
Total	102,956	87,103	84.6

→ Éligibilité selon le cycle:

On cherche la cohérence dans l'éligibilité dans tous les cycles

ESCC – Éligibilité pour appariement

		n	%*
Cohorte 1			
Province			
Terre Neuve	21,761	19,153	88.0
Ile de Prince Edouard	12,915	11,300	87.5
Nouvelle Écosse	27,600	24,672	89.4
Nouveau Brunswick	27,698	24,677	89.1
Québec	141,243	123,601	87.5
Ontario	231,932	194,805	84.0
Manitoba	41,275	35,828	86.8
Saskatchewan	42,198	36,712	87.0
Alberta	69,470	58,644	84.4
Colombie Britannique	88,604	72,986	82.4
Yukon	5,670	4,007	70.7
Territoires du Nord-Ouest	6,105	4,989	81.7
Nunavut	3,954	3,400	86.0
Sexe			
Hommes	329,478	282,580	85.8
Femmes	390,947	332,194	85.0
Groupe d'âge			
12 à 24	128,436	112,323	87.5
25 à 34	100,216	86,759	86.6
35 à 44	112,582	95,650	85.0
45 à 54	110,378	93,319	84.5
55 à 64	110,817	94,172	85.0
65 à 74	84,559	71,962	85.1
75+	73,438	60,589	82.5

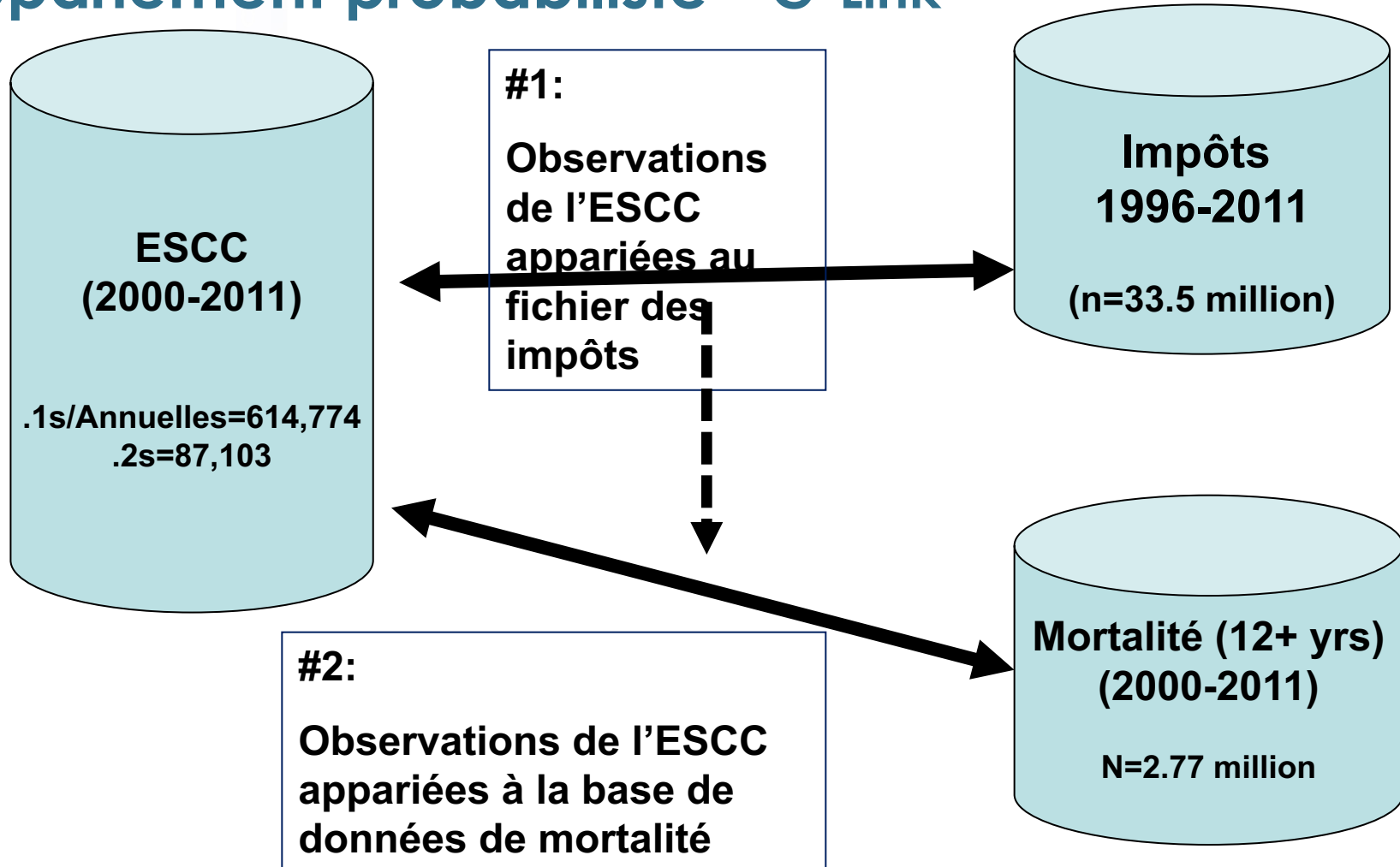
→ Éligibilité selon les variables

Éligibilité similaire dans tous les groupes, à l'exception du Yukon



Étape 2: Appariement

Appariement probabiliste – G-Link





Étape 2: Appariement

- On doit calculer le taux d'appariement par fichier, par variable
- Comme illustration, examinons les résultats concernant le fichier de mortalité



Étape 3: Validation (Fichier de mortalité)

- Révision manuelle de 1000 paires – contenant des liens et des non-liens
 - Choisies parmi 8 strates de poids d'appariements
- Information fournie
 - Variables d'appariement: noms, date de naissance, sexe, codes postaux
 - Autre information: Décès dans le fichier des impôts
- 3 personnes ont révisé les appariements
 - En cas de désaccord, la majorité l'emporte!

Validation de l'appariement de l'ESCC: Résultats concernant la mortalité

- Taux d'erreur:
 - Faux positifs = 0,04%
 - 0,04% des appariements (n=32 555 en tout) étaient erronés
 - *Très bonne performance!*
 - Faux négatifs= 3,46%
 - 3,46% des appariements ont été manqués
 - *Très bonne performance!*

Objectifs:

-????? – il n'existe pas de norme

-faible proportion de FP (<2-3%)

-faible proportion de FN (<5%)

Appariement ESCC – Mortalité

	% appariés au fichier de mortalité	
	non pondérés	pondérés
TOTAL	5.3	
	n=32 555	
2000/02 (1.1)	9.9	7.6
2003/04 (2.1)	7.6	5.2
2005/06 (3.1)	5.8	4.0
2007	4.0	2.7
2008	3.0	1.9
2009	1.9	1.1
2010	1.2	0.7
2011	0.3	0.2
TOTAL	7.4	
2002 - (1.2 - Santé mentale)	9.9	6.9
2003 (2.2 - Nutrition)	6.9	4.4
2009 (4.2 - Vieillesse en sant	6.0	3.2

On cherche:

- Un taux “raisonnable” d’appariement
- Un taux qui diminue avec le temps

Appariement ESCC – Mortalité

	ESCC (.1)		ESCC (.1, .2)	
TOTAL	5.30%		9.90%	
	non pondéré	pondéré	non pondéré	pondéré
Sexe				
Hommes	5.8	3.1	10.3	7.4
Femmes	4.9	2.6	9.5	6.5
Groupes d'âge				
12 à 24	0.3	0.2	0.5	0.3
25 à 34	0.5	0.3	0.6	0.7
35 à 44	0.9	0.6	2.0	1.7
45 à 54	2.4	1.5	4.6	3.7
55 à 64	4.7	3.1	10.9	9.5
65 à 74	11.3	7.9	20.6	17.8
75+	26.7	20.2	50.4	48.8

On cherche:

-Des tendances dans les taux d'appariements qui reproduisent les tendances observées dans le fichier de mortalité

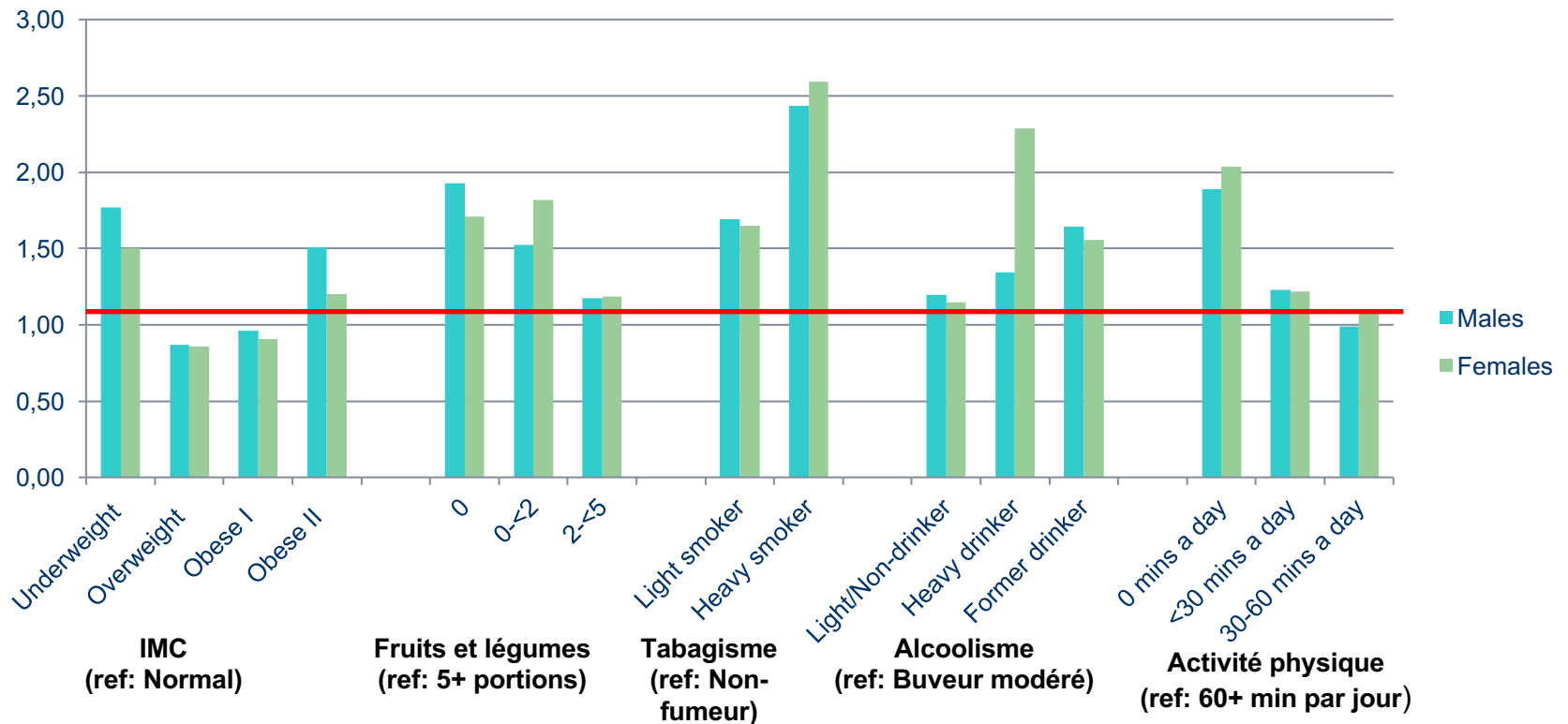
Appariement ESCC – Validation externe

- On compare la distribution des décès dans le fichier apparié à celle observée dans le fichier de mortalité seul (12 ans et +)
- Distribution pas exactement égale car la population institutionnalisée n'est pas incluse dans l'ESCC
- Les estimés pondérés sont plus représentatifs

Distribution des décès			
	ESCC .1 (apparié)		Mortalité (2000-2011)
	non pondéré	pondéré	
Province			
Terre Neuve	2.9	1.9	1.9
Ile de Prince Edouard	2.3	0.5	0.5
Nouvelle Écosse	4.9	3.8	3.6
Nouveau Brunswick	4.2	2.7	2.7
Québec	18.5	25	24.3
Ontario	31.5	36.4	37.0
Manitoba	6.8	4	4.3
Saskatchewan	7.3	3.6	3.9
Alberta	8	8.2	8.4
Colombie Britannique	12.6	13.5	13.2
Yukon	0.4	0.1	0.1
Territoires du Nord-Ouest	0.4	0.1	0.1
Nunavut	0.3	0	0.0
Sexe			
Hommes		53.8	50.5
Femmes		46.2	49.5
Groupe d'âge			
12 à 24	0.9	1.1	1.2
25 à 34	1.2	1.5	1.2
35 à 44	2.7	3.5	2.6
45 à 54	6.8	9.3	6.0
55 à 64	13.6	15.4	10.6
65 à 74	25.1	23.9	17.7
75+	49.6	45.3	60.7

Étape charnière: Validation externe et Recherche

Rapports de cotes ajustés selon l'âge et le sexe de facteurs de risque associés à la mortalité en 5 années de suivi.





Récapitulation: Appariement avec ESCC

- Complété: Données ESCC 2000-2011 appariées à:
 - Congés d'hospitalisation, 1996/1997 à 2012/13
 - Mortalité, 2000 à 2011
 - [Sanmartin et al. *Linking the Canadian Community Health Survey and the Canadian Mortality Database: An enhanced data source for the study of mortality. Health Reports December 2016*](#)
- Date de rétention date: 31 décembre 2020 (option pour extension)
- Accès aux CDR:
 - Disponibilité – 2017/18

Autres appariements

- 1) *Recensement 1991 – Mortalité – Cancer: dans les CDR*
- 2) *Santé et environnement – dans les CDR*
- 3) *Recensement - Naissances (1996, 2006) – dans les CDR*
- 4) *Recensement 2006 (formulaire long) avec congés d'hospitalisation –
Projet pilote (CDR)*
- 5) *Registre du Cancer et congés d'hospitalisation*
- 6) *Fichier d'immigration 1980-2006 et congés d'hospitalisation*



Articles et résultats

Articles de validation

N° 82-003-X au catalogue
ISSN 1209-1375

Rapports sur la santé

Utilisation des numéros d'assurance-maladie personnels pour coupler le Registre canadien du cancer et la Base de données sur les congés des patients

par Dianne Zakaria, Richard Trudeau, Claudia Sanmartin, Patricia Murison, Gisèle Carrière, Maureen MacIntyre, Donna Turner, Brandon Wagar, Mary Jane King, Kim Vriends, Ryan Woods, Gina Lockwood et Rabiâ Louchini

Date de diffusion : le 17 juin 2015



Articles de validation

N° 82-003-X au catalogue
ISSN 1209-1375

Rapports sur la santé

Couplage des données du Recensement de 2006 et des données hospitalières au Canada

par Michelle Rotermann, Claudia Sanmartin, Richard Trudeau et Hélène St-Jean

Date de diffusion : le 21 octobre 2015





Articles de validation

N° 11-633-X au catalogue — N° 002
ISSN 2371-3437
ISBN 978-0-660-05810-8

Études analytiques : méthodes et références

Couplage du fichier canadien d'établissement des immigrants aux données hospitalières : une nouvelle source de données pour la recherche sur la santé des immigrants

par Claudia Sanmartin, Edward Ng, James Brennan,
Scott McLeish, Richard Trudeau et Doug Manuel

Date de diffusion : le 17 août 2016



Articles de validation

N° 82-003-X au catalogue
ISSN 1209-1375

Rapports sur la santé

Couplage de l'Enquête sur la santé dans les collectivités canadiennes et de la Base canadienne de données sur la mortalité : une source de données améliorée pour l'étude de la mortalité

par Claudia Sanmartin, Yves Decady, Richard Trudeau, Abel Dasyilva, Michael Tjepkema, Philippe Finès, Rick Burnett, Nancy Ross et Douglas G. Manuel

Date de diffusion : le 21 décembre 2016



Articles de résultats

Rapports sur la santé

Hospitalisation toutes causes confondues et liée aux maladies de l'appareil circulatoire, selon le statut des générations : résultats d'un couplage de données

par Edward Ng, Claudia Sanmartin, Jack V. Tu et Douglas G. Manuel

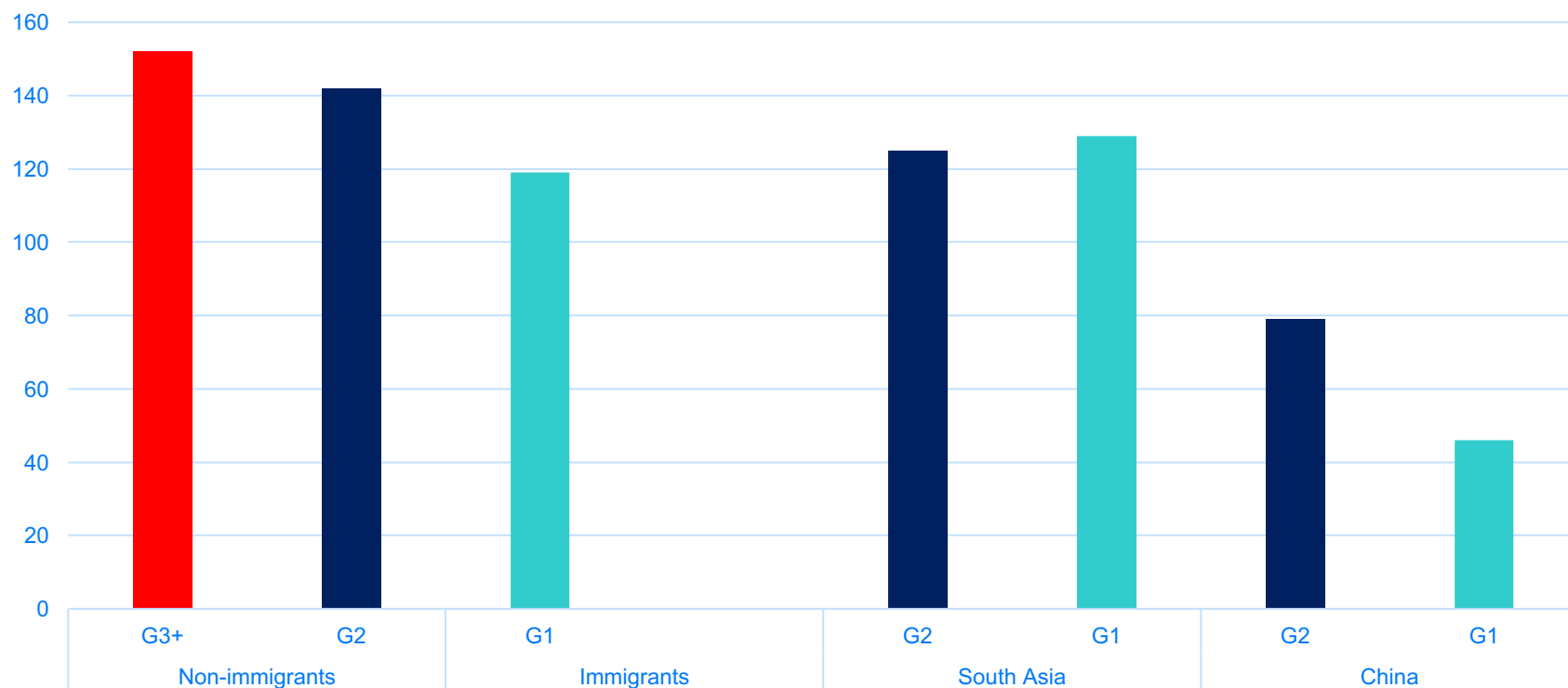
Date de diffusion : le 21 octobre 2015



Résultats de recherche

(Ng E, Sanmartin C et al. [Rapports sur la santé](#), October 2015)

Taux standardisé d'hospitalisations pour maladies du système circulatoire (per 10,000 pop) par statut d'immigrant et génération, Canada (sans Québec), 2006-2009



Source: Recensement 2006, Congés d'hospitalisation 2006/2007, 2007/2008, 2008/2009

Articles de résultats

N° 82-003-X au catalogue
ISSN 1209-1375

Rapports sur la santé

Hospitalisation en soins de courte durée, selon l'identité autochtone, Canada, 2006 à 2008 inclusivement

par Gisèle Carrière, Evelyne Bougie, Dafna Kohen, Michelle Rotermann et Claudia Sanmartin

Date de diffusion : le 17 août 2016

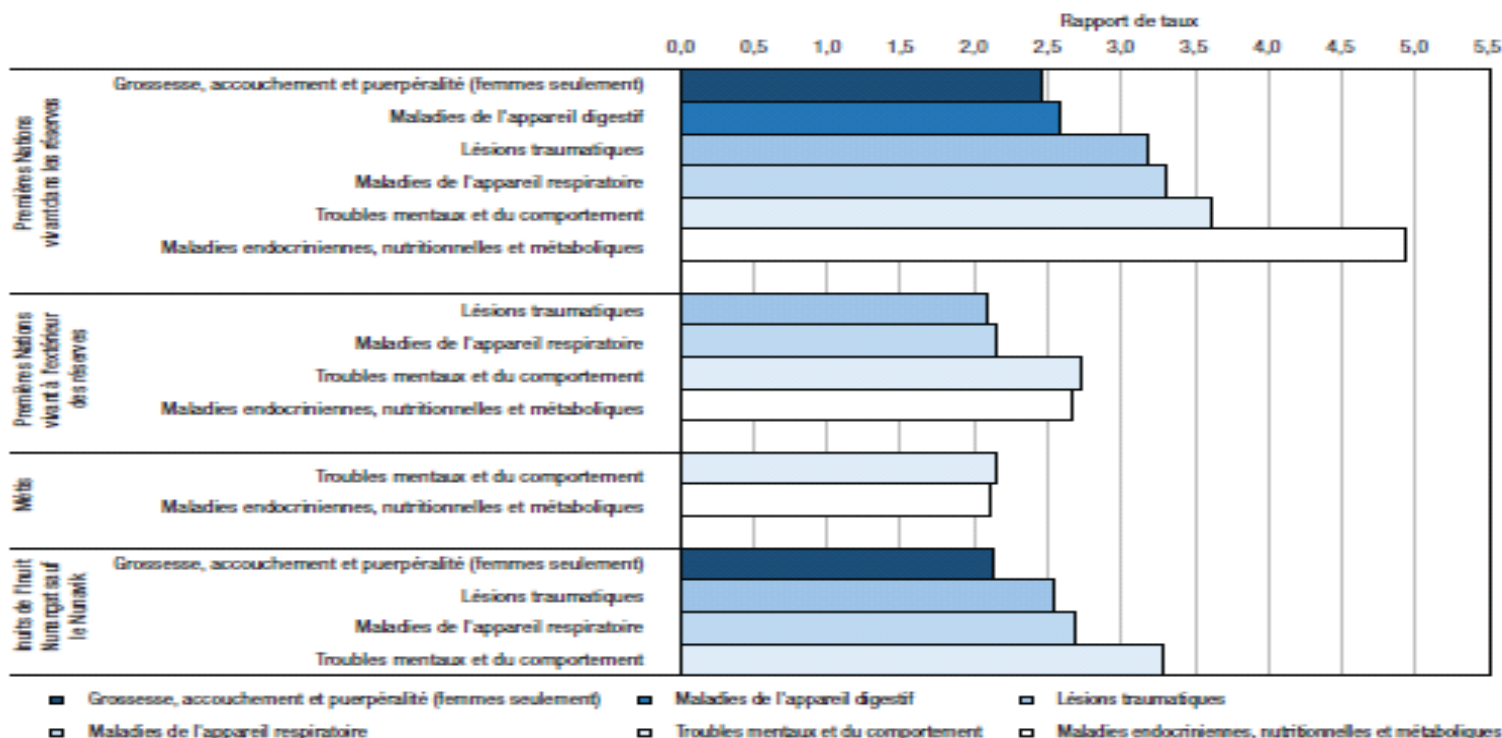




Résultats de recherche

(Carriere G, Bougie E et al. Rapports sur la santé): Hospitalisations pour soins de courte durée

Figure 1
Rapports de taux de 2,0 ou plus[†] pour les taux d'hospitalisation en soins de courte durée normalisés selon l'âge[‡] pour 100 000 habitants, selon l'identité autochtone et la cause, cohorte d'étude définie par suite du couplage Recensement de 2006—Base de données sur les congés des patients, Canada sauf le Québec, 2006-2007 à 2008-2009



[†] le groupe de référence est la population non autochtone

[‡] normalisés en fonction de la structure par âge de la population autochtone d'après le Recensement de 2006

Source : Données résultant du couplage Recensement de 2006—Base de données sur les congés des patients de 2006-2007 à 2008-2009.

Source: Recensement 2006, Congés d'hospitalisation 2006/2007, 2007/2008, 2008/2009



Articles de résultats

N° 11-633-X au catalogue — N° 003
ISSN 2371-3437
ISBN 978-0-660-06458-1

Études analytiques : méthodes et références

Cohorte du Recensement de 2001 couplée avec des données fiscales et des données sur la mortalité : un suivi de 10 ans

par Lauren Pinault, Philippe Finès, Félix Labrecque-Synnott,
Abdelnasser Saidi, et Michael Tjepkema

Date de diffusion : le 26 octobre 2016



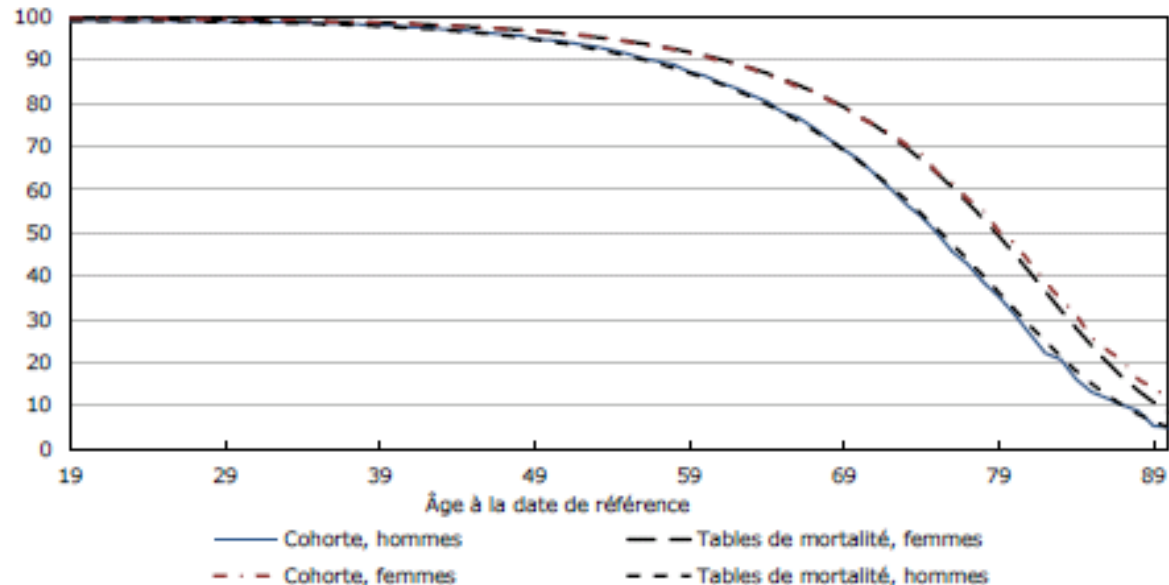


Résultats de recherche

(Pinault L. et al. Rapports sur la santé): Cohorte du Recensement du Canada de 2001 couplée avec des données fiscales et des données sur la mortalité : un suivi de 10 ans

Graphique 1
Pourcentage de répondants ayant survécu 10,6 ans, selon l'âge et le sexe, suivi de la mortalité de 2001 à 2011, comparativement aux tables de mortalité du Canada (CANSIM) pour 2005 à 2007

pourcentage de survie



Sources : Statistique Canada, cohorte constituée par le couplage des données du Recensement du Canada de 2001, des données fiscales et des données sur la mortalité (définie à partir du Recensement de la population de 2001 et de la Base de données combinées sur la mortalité de 2014), et calculs des auteurs d'après des données de Statistique Canada, 2013, *Tables de mortalité, Canada, provinces et territoires, 2005 à 2007*.

Articles de résultats

N° 11-633-X au catalogue — N° 006
ISSN 2371-3437
ISBN 978-0-660-07656-0

Études analytiques : méthodes et références

Imputation de codes postaux en vue de l'analyse de variables écologiques dans les cohortes longitudinales : exposition aux matières particulaires dans la base de données Cohorte santé et environnement du Recensement du Canada

par Philippe Finès, Lauren Pinault, et Michael Tjepkema

Date de diffusion : le 13 mars 2017



Résultats de recherche

(Finès P. et al. Études analytiques):
Imputation de codes postaux

Tableau 3
Répartition de la durée des vides dans CSERCan après retrait des codes postaux non résidentiels

Durée du vide en années	Répartition	
	nombre	pourcentage
1	972 655	40,0
2	360 670	14,8
3	204 310	8,4
4	145 625	6,0
5	125 000	5,1
6	111 065	4,6
7	74 300	3,1
8	44 515	1,8
9	42 860	1,8
10	39 975	1,6
11	34 040	1,4
12	35 225	1,4
13	33 810	1,4
14	32 710	1,3
15	32 600	1,3
16	31 335	1,3
17	29 220	1,2
18	27 670	1,1
19	26 285	1,1
20	19 780	0,8
21	7 815	0,3
22	220	0,0
23	115	0,0
24	75	0,0
25	80	0,0
26	50	0,0
27	25	0,0
Total	2 431 995	100,0

Notes : Le total peut ne pas être égal à la somme des nombres des lignes précédentes en raison des règles de confidentialité. De plus, la somme des pourcentages n'est pas égale à 100,0 % en raison de l'arrondissement.
Source : Statistique Canada, Cohorte santé et environnement du Recensement du Canada (CSERCan) de 1991 couplée au Fichier de données fiscales sommaires historiques pour 1984 à 2011.



Adresses utiles

- Appariement à StatCan - <http://www.statcan.gc.ca/eng/record/gen>
- Services d'appariement – l'Environnement de couplage de données sociales (ECDS) <http://www.statcan.gc.ca/eng/sdle/index>
- Centres de données de recherche - <http://www.statcan.gc.ca/fra/cdr/index>
- Données disponibles dans les centres de données de recherche - <http://www.statcan.gc.ca/fra/cdr/donnees>
- Démarches d'accès: <http://www.statcan.gc.ca/fra/cdr/process>
- Pour les chercheurs du Québec, les CDR sont en réseau sous l'égide du CIQSS: le Centre interuniversitaire québécois de statistiques sociales.
- <https://www.ciqss.org/> et courriel: Accès CIQSS <acces@ciqss.org> (Catherine Blanchard-Gougeon)

Conclusions

- Beaucoup d'activités d'appariement
- Source importante de “nouvelles” données pour la recherche
- Statistique Canada est leader dans les méthodes et procédures
- Accès aux CDR en cours



Merci!

- Philippe Finès
- philippe.fines@canada.ca