

# Qui dit vrai : participants ou bases de données médico-administratives ?

Yves Payette<sup>1</sup>, Cristiano Moura<sup>2,4</sup>, Catherine Boileau<sup>1</sup>, Catherine Labbé<sup>1</sup>, Sébastien Jacquemont<sup>3</sup>, Sasha Bernatsky<sup>2,4</sup>, Nolwenn Noisel<sup>1,5</sup>

<sup>1</sup>CARTaGENE, Centre de recherche CHU Sainte-Justine, Montréal, Canada; <sup>2</sup>Division of Clinical Epidemiology, McGill University, Montréal; <sup>3</sup>Département de pédiatrie, Faculté de médecine, Université de Montréal, CHU Sainte-Justine, Montréal; <sup>4</sup>Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal; <sup>5</sup>École de santé publique, Université de Montréal, Montréal.

## CONTEXTE ET OBJECTIFS

De nombreuses études épidémiologiques utilisent des données de questionnaires auto-rapportées pour évaluer l'exposition, les effets de santé, la prévalence des maladies chroniques ou la prise de médication. Cependant, la fiabilité et la précision de ces données auto-rapportées varient selon le statut socioéconomique ou le type de maladie ciblée<sup>1</sup>.

Pour cette raison, les données médico-administratives sont souvent utilisées pour valider des données auto-rapportées<sup>2,3</sup>, même si celles-ci ne sont pas collectées initialement pour des fins de surveillance.

**L'utilisation de données médico-administrative soulève également des défis puisque des erreurs de classification sont aussi possibles.**

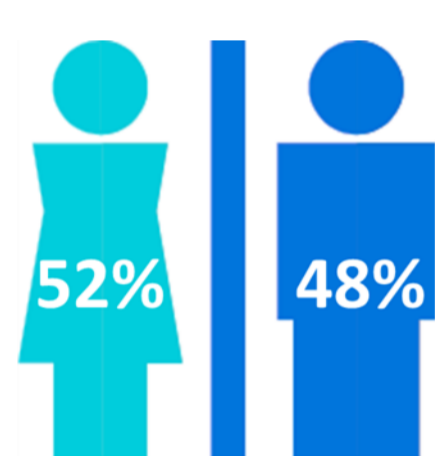
Les objectifs de cette étude sont : i) d'analyser la concordance des sources de données (auto-rapportées vs. médico-administratives) relativement aux diagnostics de maladies, et ii) d'identifier les facteurs influençant la concordance entre les sources de données.

## MÉTHODE

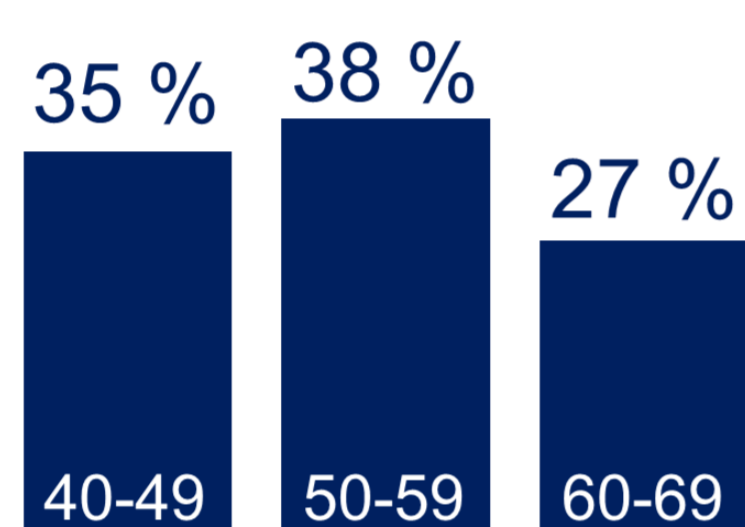
- L'étude a été réalisée à l'aide des bases de données médico-administratives (RAMQ, ISQ, INSPQ) et des données de la cohorte populationnelle CARTaGENE.
- CARTaGENE (CaG) constitue une ressource publique unique avec 43 000 participants représentatifs de la population âgée de 40 à 69 au Québec. Les données collectées par CaG regroupent des questionnaires de santé, de nutrition, d'environnement et d'habitudes de vie, des mesures physiques et des échantillons biologiques.
- L'accès aux données médico-administratives (RAMQ, ISQ, INSPQ) des participants de CaG, à partir de 1998 et mises à jour annuellement, permet de reconstruire et comprendre les trajectoires de comorbidité et de mieux prédire les tendances futures.
- L'analyse de la concordance entre les données auto-rapportées des participants de CaG et les données médico-administratives (diagnostic des maladies à partir des CIM-9 et CIM-10) a été réalisée par le calcul d'indicateurs de concordance (sensibilité, spécificité, valeur prédictive positive et kappa) pour 19 maladies chroniques et 12 types de cancers.
- Des modèles de régression logistiques ont été développés afin d'identifier les facteurs influençant la concordance.

## CARACTÉRISTIQUES DES PARTICIPANTS

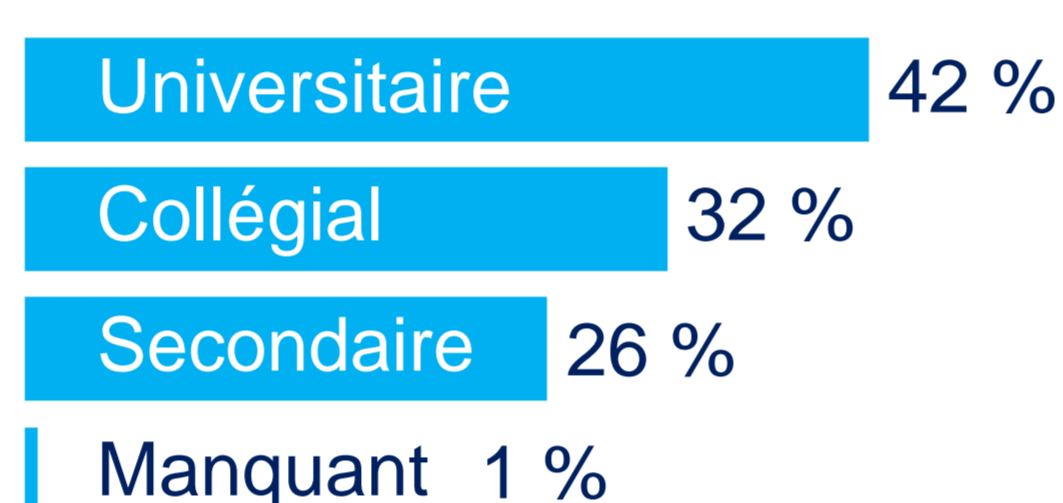
### Distribution des sexes



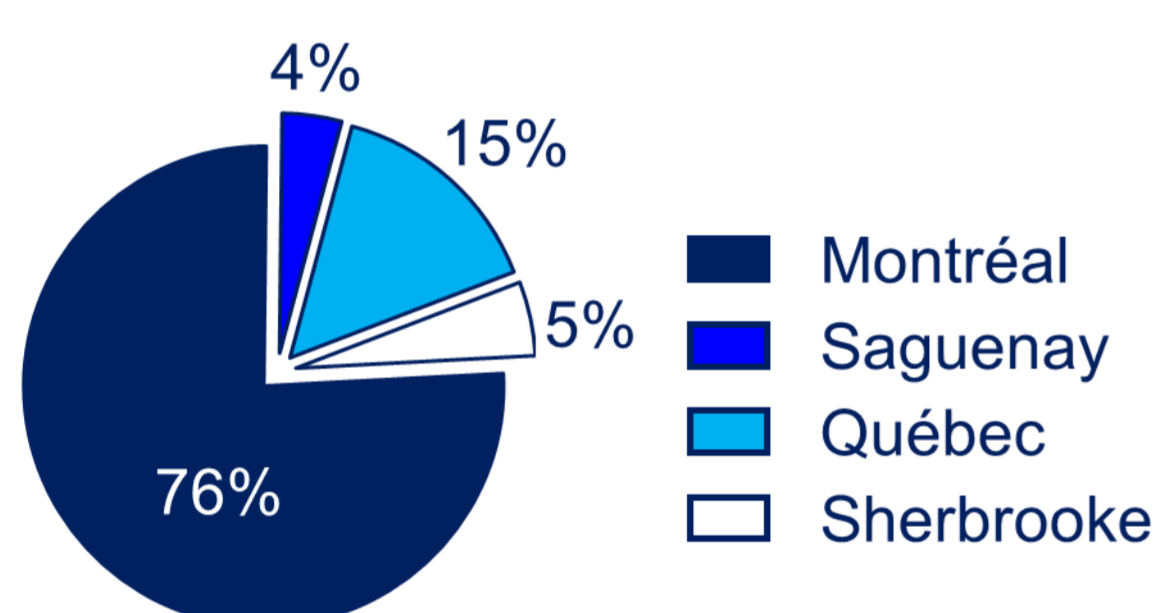
### Distribution des âges



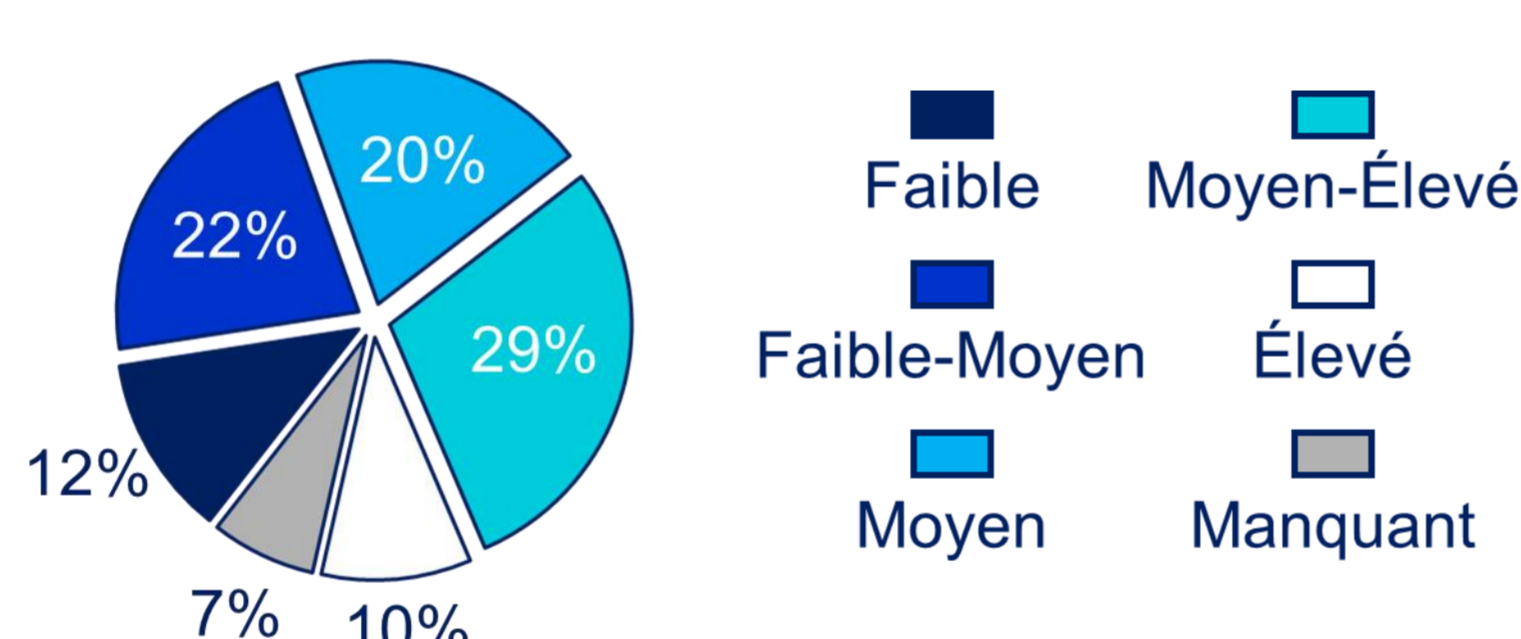
### Niveau d'éducation



### Représentation des régions



### Quintiles de revenu



Utilisation des soins de santé dans les 3 dernières années	N	%
< 20 réclamations	8837	44,2
≥ 20 réclamations	11159	55,8
Indice de comorbidité de Charlson <sup>4</sup> (CCI)	N	%
0	15159	75,8
1	2802	14,0
2	1131	5,7
3 à 12	904	4,5

## FRÉQUENCES DE MORBIDITÉS

Maladies et cancers étudiés	Fréquences (%)		Sensibilité	Spécificité	VPP	Kappa
	CaG	BDMA				
<b>Maladies</b>						
Hypertension *	25,0	22,8	73,83	94,15	80,79	0,70
Ostéoarthrite	16,1	20,5	38,48	89,70	49,05	0,31
Dépression	16,7	18,4	51,16	91,59	57,83	0,45
Diabète	7,6	8,0	85,06	99,15	89,74	0,86
MPOC	5,7	7,6	22,46	95,69	30,06	0,21
Hypothyroïdisme	7,1	7,6	61,52	97,40	66,10	0,61
Asthme *	13,0	6,4	78,15	91,43	38,31	0,47
IRC	0,3	3,3	5,09	99,91	72,13	0,09
AVC	1,7	2,1	41,36	99,18	51,67	0,45
SCI *	4,5	1,8	46,96	96,31	19,10	0,25
IM *	2,8	1,7	88,62	98,66	52,95	0,66
Maladie de Crohn	0,6	1,4	38,01	99,91	85,83	0,52
AR *	2,9	0,9	71,10	97,74	21,81	0,33
Schizophrénie	0,4	0,8	35,37	99,91	77,33	0,48
SP	0,4	0,7	56,49	99,96	90,24	0,69
Épilepsie	0,6	0,6	59,63	99,76	58,04	0,59
Cirrhose	0,3	0,4	31,71	99,85	47,27	0,38
Maladie de Parkinson	0,1	0,3	38,10	100,00	100,00	0,55
LED	0,2	0,3	35,94	99,84	41,82	0,39
<b>Cancers</b>						
Sein	1,8	1,9	86,28	99,88	93,43	0,90
Prostate	0,9	1,3	67,60	99,95	94,94	0,79
Mélanome *	1,7	0,8	45,28	98,68	21,62	0,29
Colon	0,4	0,7	43,57	99,89	73,49	0,55
LNH	0,2	0,5	36,79	99,97	88,64	0,52
Poumon	0,3	0,5	47,37	99,96	84,91	0,61
Col utérin *	0,7	0,4	28,00	99,45	16,03	0,20
Vessie	0,2	0,3	58,82	99,98	93,02	0,72
Thyroïde	0,2	0,3	71,67	99,94	79,63	0,75
Rein	0,2	0,3	57,41	99,98	91,18	0,70
Rectum	0,1	0,3	24,00	99,98	75,00	0,36
Utérus *	0,4	0,2	47,37	99,70	23,08	0,31

CaG : Diagnostique auto-rapporté dans la cohorte CaG; BDMA : Bases de données médico-administratives; VPP : valeur prédictive positive; MPOC : Maladie Pulmonaire Obstructive Chronique; IRC : Insuffisance Rénale Chronique; AVC : Accident Vasculaire Cérébral; SCI : Syndrome du Côlon Irritable; IM : Infection du Myocarde; AR : Arthrite Rhumatoïde; SP : Sclérose en Plaques; LED : Lupus Érythémateux Disséminé; LNH : Lymphome Non-Hodgkinien. Aucun cas de cancer de la trachée identifié dans les bases de données médico-administratives.  
\* Fréquences inférieures au sein des BDMA par rapport à celles auto-rapportées par les participants de CaG.

## RÉSULTATS

- Pour la majorité des maladies et des cancers étudiés (12/19 et 8/12, respectivement), la concordance est modérée à bonne (kappa > 0,41).
- La prévalence est généralement plus élevée pour les données médico-administratives que pour les données auto-rapportées (13/19 et 9/12 respectivement), en particulier pour l'insuffisance rénale chronique (3,30% vs. 0,3%).
- Cependant, pour des maladies telles que l'asthme, le syndrome du côlon irritable et l'arthrite rhumatoïde, les données médico-administratives semblent sous-estimer les prévalences (6,40% vs. 13,0%, 1,70% vs. 2,80% et 0,90% vs. 2,90%, respectivement pour les données médico-administratives et les données auto-rapportées).
- Les analyses de régression logistiques montrent que la concordance est associée avec l'âge, le sexe (femme vs. homme), le revenu et de façon inverse avec l'utilisation des services de santé et la comorbidité (p<0,05).

## CONCLUSION

- Bien qu'il existe des différences entre les données de questionnaires auto-rapportées et les données médico-administratives en ce qui concerne le diagnostic de maladies, les données auto-rapportées démontrent tout autant leur validité scientifique.
- Les raisons expliquant ces différences sont diverses et incluent différents biais (biais de mémoire, de désirabilité, de classification, etc.).
- Cette étude suggère l'importance de bien connaître les biais associés à chaque source de données afin de faire le meilleur choix en fonction du type d'étude réalisée.

## RÉFÉRENCES

- Short, M. E. et al. 2009 J Occup Environ Med 51(7): 786-796.
- Okura, Y. et al. 2004 J Clin Epidemiol 57(10): 1096-1103.
- Muggah, E. et al. 2013 BMC Public Health 13: 16.
- Crump, C. et al. 2017 Am J Prev Med 52(3): 353-361.