

N° 11-633-X au catalogue — N° 026
ISSN 2371-3437
ISBN 978-0-660-35653-2

Études analytiques : méthodes et références

Élaboration de catégories utiles permettant de distinguer les niveaux d'éloignement au Canada

par Rajendra Subedi, Shirin Roshanafshar et T. Lawson Greenberg

Date de diffusion : le 11 août 2020

 Statistique
Canada

Statistics
Canada

Canada 

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

Programme des services de dépôt

- | | |
|-----------------------------|----------------|
| • Service de renseignements | 1-800-635-7943 |
| • Télécopieur | 1-800-565-7757 |

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté la Reine du chef du Canada, représentée par le ministre de l'Industrie 2020

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Élaboration de catégories utiles permettant de distinguer les niveaux d'éloignement au Canada

par

Rajendra Subedi, Shirin Roshanafshar et T. Lawson Greenberg

Centre de données sur la santé de la population

Statistique Canada

11-633-X — 2020002 — N° 026

ISSN 2371-3437

ISBN 978-0-660-35653-2

Août 2020

Études analytiques : méthodes et références

Les documents de cette série traitent des méthodes utilisées pour produire des données qui seront employées pour effectuer des études analytiques à Statistique Canada sur l'économie, la santé et la société. Ils ont pour but de renseigner les lecteurs sur les méthodes statistiques, les normes et les définitions utilisées pour élaborer des bases de données à des fins de recherche. Tous les documents de la série ont fait l'objet d'un examen par les pairs et d'une révision institutionnelle, afin de veiller à ce qu'ils soient conformes au mandat de Statistique Canada et qu'ils respectent les normes généralement reconnues régissant les bonnes pratiques professionnelles.

Les documents peuvent être téléchargés gratuitement de www.statcan.gc.ca.

Table des matières

Remerciements	5
Résumé	6
1 Introduction	7
2 Données et méthodes	9
3 Résultats	12
3.1 Classification manuelle des valeurs de l'indice d'éloignement en fonction des seuils naturels, des chiffres de population et de la répartition des subdivisions de recensement	12
3.2 Classification par intervalles égaux fondée sur la valeur de l'indice d'éloignement	14
3.3 Classification par quantiles en fonction du nombre de subdivisions de recensement	15
3.4 Classification en fonction des seuils naturels de Jenks	17
3.5 Classification des valeurs de l'indice d'éloignement par K-moyennes.....	19
4 Conclusions et recommandations	21
Bibliographie	23

Remerciements

Nous tenons à remercier Erin Pichora, de l'Institut canadien d'information sur la santé, ainsi qu'Alessandro Alasia et François Sergerie, de Statistique Canada, pour leurs commentaires et leurs suggestions utiles à titre d'examineurs officiels de ce document.

Résumé

Les concepts de région urbaine et de région rurale font l'objet de nombreux débats et varient selon la composition géopolitique et sociodémographique de chaque pays. Au Canada, les classifications des centres de population et des secteurs statistiques sont largement utilisées pour distinguer les collectivités urbaines des collectivités rurales. Toutefois, aucune de ces classifications ne ventile précisément les collectivités canadiennes en régions urbaines, rurales et éloignées. Pour combler cette lacune, un groupe de chercheurs de Statistique Canada a mis au point un outil appelé l'« indice d'éloignement » pour mesurer l'éloignement relatif des collectivités canadiennes. La présente étude met à profit cet indice d'éloignement, qui est un indice continu, pour aborder la façon dont les collectivités canadiennes peuvent être classées en cinq catégories distinctes de régions géographiques éloignées. Lorsqu'il est utilisé adéquatement, l'indice d'éloignement peut aider à distinguer les collectivités urbaines, rurales et éloignées du Canada, tout en protégeant la vie privée et la confidentialité des citoyens. Les auteurs de la présente étude y examinent cinq approches méthodologiques et en recommandent trois.

Mots clés : classification, éloignement, grappe, rural, seuil naturel, urbain

1 Introduction

Les concepts de région urbaine et de région rurale font l'objet de nombreux débats partout dans le monde. Les définitions de ces concepts varient selon la situation géopolitique et la composition sociodémographique de chaque pays. Certains chercheurs ont défini les régions rurales comme des régions géographiques caractérisées par des structures démographiques et des modèles d'établissement qui leur sont propres, des populations isolées, de longues distances de navettage, un approvisionnement restreint en biens et en services, ainsi qu'une économie majoritairement fondée sur l'agriculture et les ressources naturelles (Hart, Lishner et Larson, 2005; Johnson et Johnson, 2015; du Plessis et coll., 2002). Cependant, il n'existe pas de définition des termes « région urbaine » et « région rurale » ou de classification des régions urbaines, rurales et éloignées qui soient universellement acceptées.

En général, les pays ont défini les régions rurales et éloignées en fonction de leur accessibilité géographique et leur densité de population. Par exemple, aux États-Unis, les comtés et les entités équivalentes dont la population est inférieure à 10 000 habitants sont classés comme des comtés non centraux et sont généralement considérés comme des régions rurales (National Center for Health Statistics [NCHS], 2014). Au Royaume-Uni, selon la classification des régions rurales et urbaines de 2011, les régions qui comptent une population de 10 000 habitants ou plus sont qualifiées d'urbaines, et les régions restantes, de rurales (Bibby et Brindley, 2013). Cependant, les critères appliqués aux États-Unis et au Royaume-Uni ne sont pas applicables dans le contexte canadien, puisque la densité de la population au Canada est relativement faible.

En Australie, il existe trois mesures de l'accessibilité et de l'éloignement, soit la classification géographique des régions rurales, éloignées et métropolitaines (ministère des Industries primaires et de l'Énergie et ministère des Services humains et de la Santé, 1994), l'indice d'accessibilité et d'éloignement de l'Australie (ministère de la Santé et des Soins aux aînés, 2001) et la classification géographique type de l'Australie (Institut australien pour la santé et le bien-être, 2004). Ces trois classifications divisent les régions géographiques australiennes en régions urbaines, rurales et éloignées en fonction de l'accessibilité de ces régions à une gamme de services.

En Europe, cependant, une grande partie du territoire ne peut être catégorisée selon la dichotomie opposant les régions urbaines aux régions rurales, mais se situe plutôt quelque part entre les deux, en raison surtout des régions rurales densément peuplées (Wandl et coll., 2014). Par conséquent, les organismes statistiques nationaux des pays européens utilisent des définitions différentes des régions urbaines et rurales. Cela fait en sorte que les régions urbaines et rurales varient considérablement d'un pays à l'autre (Peen et coll., 2010).

Dans le contexte canadien, la classification des régions en régions urbaines et rurales la plus courante est le concept des centres de population (CTRPOP)¹ (auparavant appelés « régions urbaines »). Statistique Canada utilise les CTRPOP depuis 2011, le plus souvent dans le cadre d'études démographiques au Canada. Toutes les collectivités à l'extérieur des CTRPOP sont considérées comme étant rurales. Bien que les CTRPOP soient classés en trois groupes selon qu'ils sont petits (population de 1 000 à 29 999 habitants), moyens (population de 30 000 à 99 999 habitants) ou grands (population de 100 000 habitants ou plus), ils ne tiennent pas compte des différences entre les petits centres urbains et les régions métropolitaines de plus d'un million d'habitants.

Le concept de CTRPOP classe toutes les collectivités de moins de 1 000 habitants — ou d'une densité de moins de 400 personnes au kilomètre carré — comme étant rurales, malgré le fait que certaines de ces régions bénéficient de leur proximité avec de grands centres urbains. Par conséquent, le concept de CTRPOP tient compte de la taille et de la densité de la population, mais ne tient pas compte de la proximité avec de grands centres urbains qui fournissent des biens et des services aux petites villes. De plus, l'un des principaux inconvénients de l'utilisation des CTRPOP dans la classification des régions selon qu'elles sont urbaines ou rurales est qu'ils sont fondés sur les îlots de diffusion, auxquels un code postal ne peut être attribué de façon précise, et ce, pour environ 25 % de la population (à savoir, principalement, les personnes qui vivent en banlieue urbaine et dans les régions rurales) (Statistique Canada, 2017).

Statistique Canada a élaboré l'indice d'éloignement en 2017 en complément aux méthodes de classification géographique existantes (voir Alasia et coll., 2017). Bien que la présente étude vise à élaborer des catégories distinctes de l'indice d'éloignement, les auteurs abordent également d'autres méthodes de classification

1. Régions ayant une population d'au moins 1 000 habitants et une densité de population d'au moins 400 habitants au kilomètre carré.

géographique afin de présenter des renseignements sur le contexte et les limites des méthodes existantes. L'indice d'éloignement fournit des valeurs relatives d'éloignement pour presque toutes les subdivisions de recensement (SDR) du Canada. Il n'est donc pas possible d'utiliser les CTRPOP pour catégoriser les valeurs de l'indice d'éloignement, puisque les limites des CTRPOP ne correspondent pas exactement aux limites des SDR. Par exemple, cinq CTRPOP chevauchent les frontières provinciales, mais toutes les SDR suivent les frontières provinciales ou territoriales (Statistique Canada, 2016a).

Une autre méthode de classification géographique utilisée au Canada est la Classification des secteurs statistiques (CSS), qui regroupe les SDR selon qu'elles font partie d'une région métropolitaine de recensement (RMR), d'une agglomération de recensement (AR), d'une zone d'influence métropolitaine (ZIM) ou d'un territoire (Statistique Canada, 2016b). Une ZIM est ensuite divisée en quatre catégories selon son degré d'influence (forte, modérée, faible ou aucune influence), selon le pourcentage de la population qui se déplace pour se rendre au travail dans une ou plusieurs RMR ou AR. La CSS repose sur l'hypothèse selon laquelle les gens qui vivent près d'une RMR ou d'une AR ont davantage tendance à se déplacer pour se rendre au travail. La CSS est le plus souvent utilisée pour faire rapport sur les indicateurs de la santé ou d'autres données lorsque l'information sur le lieu de résidence se limite au code postal. Toutefois, la CSS ne mesure pas précisément l'accès aux biens et services offerts à l'intérieur ou à proximité d'une collectivité. De plus, la CSS regroupe l'ensemble des SDR dans les territoires au sein d'une même catégorie, même si certaines régions sont plus accessibles que d'autres.

Le concept d'éloignement est particulièrement pertinent dans le contexte de l'étude des caractéristiques socioéconomiques et de l'état de santé de la population, étant donné que l'information géographique contenue dans les données sur la santé se limite souvent au code postal résidentiel. Des études ont montré que les populations rurales et éloignées sont caractérisées par un moins bon état de santé, une mortalité plus élevée, une espérance de vie plus courte et des besoins en soins de santé non comblés plus élevés (Eckert, Taylor et Wilkinson, 2004; Mitura et Bollman, 2003). Par exemple, une analyse de l'Institut canadien d'information sur la santé (ICIS) donne à penser que les taux d'hospitalisation et de mortalité liés à l'asthme, aux maladies pulmonaires obstructives chroniques, au diabète, à l'hypertension artérielle et aux maladies cardiaques sont disproportionnellement plus élevés chez les Canadiens qui habitent en région rurale que chez ceux qui habitent en région urbaine, même lorsque l'on tient compte des taux de prévalence des maladies plus élevés (ICIS, 2012). Dans une analyse antérieure, l'ICIS a examiné l'état de santé des Canadiens selon le degré de ruralité au moyen des ZIM (Desmeules et Pong, 2006). Toutefois, dans cette étude, l'ICIS a utilisé les ZIM modérées et faibles ainsi que les territoires comme régions éloignées.

Le nouvel indice d'éloignement pourrait constituer une solution de rechange plus fiable aux concepts traditionnels de classification géographique des CTRPOP et de la CSS pour le Canada, puisqu'il attribue une valeur d'éloignement relatif à chacune des SDR en fonction de leur proximité avec les agglomérations.

En 2018, l'indice d'éloignement a été mis à jour en fonction des régions géographiques et des populations des SDR de 2016. Il a permis de mesurer l'éloignement relatif de chaque SDR sur une échelle normalisée de 0 à 1, où 0 est la région la plus accessible (urbaine) et 1 est la région la moins accessible (éloignée). L'un des avantages de l'indice d'éloignement est la possibilité de le mettre à jour après chaque recensement de la population, qui a lieu tous les cinq ans.

Bien que l'indice d'éloignement soit un concept utile pour comprendre l'éloignement relatif de chaque SDR, son échelle continue est moins utile pour la diffusion de renseignements statistiques, à moins que les SDR soient regroupées en quelques catégories avec des indices d'éloignement semblables. Un regroupement significatif des valeurs de l'indice d'éloignement aiderait à diffuser des indicateurs socioéconomiques et de santé, sans risquer de compromettre la vie privée et les renseignements confidentiels des personnes.

L'objectif du présent document méthodologique est d'étudier différentes approches pour catégoriser les valeurs de l'indice d'éloignement continu des SDR canadiennes en groupes significatifs, et de recommander quelques classifications aux fins d'utilisation future. Ces classifications pourront servir de complément aux classifications traditionnelles des régions selon qu'elles sont urbaines ou rurales. Ce document aborde cinq approches méthodologiques pour classer l'éloignement en cinq catégories distinctes, à savoir la classification manuelle, la classification par intervalles égaux, la classification par quantiles, la classification en fonction des seuils naturels de Jenks et la classification par K-moyennes. Chaque méthode sera appliquée de sorte à regrouper ensemble les SDR dont les valeurs de l'indice d'éloignement sont semblables afin de former des catégories d'éloignement significatives.

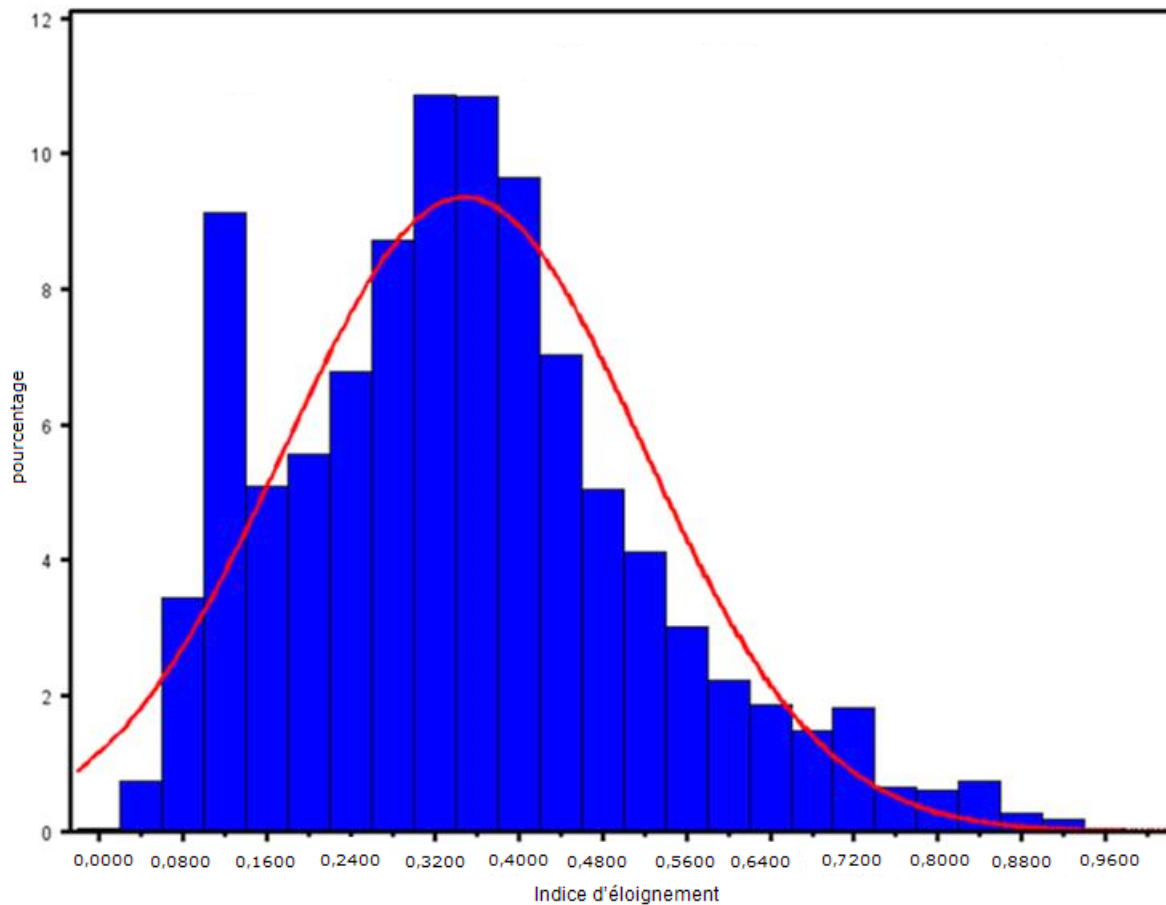
2 Données et méthodes

L'ensemble de données utilisé dans cette classification est composé des valeurs de l'indice d'éloignement mises à jour en fonction des régions géographiques et des populations des SDR de 2016 (Alasia et coll., 2017). L'indice d'éloignement découle lui-même de la combinaison de données de sources statistiques officielles, comme le Recensement de la population, et de données de sources statistiques non officielles, comme l'interface de programmation d'application (IPA) Google Maps. L'unité géographique utilisée aux fins d'analyse dans l'indice d'éloignement est la SDR. Pour calculer la valeur de l'indice d'éloignement, on a combiné la couche géographique correspondant aux SDR et celle correspondant aux CTRPOP. On a déterminé la valeur de l'indice d'éloignement de chaque SDR en fonction de la proximité relative de la SDR (mesurée en coût de déplacement²) avec tous les CTRPOP environnants dans un rayon de 200 kilomètres. La taille de la population de chaque CTRPOP a été utilisée comme indicateur de la disponibilité des services. Le calcul de l'indice d'éloignement tient compte de tous les CTRPOP qui pourraient être des emplacements potentiels de biens, de services et d'activités économiques pour la SDR de référence. L'indice d'éloignement mis à jour comprenait des valeurs d'indice pour toutes les SDR au Canada qui ont déclaré une population en 2016 ou qui étaient rattachées au principal réseau de transport (5 125 SDR sur 5 162).

La figure suivante illustre la répartition des valeurs de l'indice d'éloignement des SDR en 2016. Comme le montre la figure 1, de nombreuses SDR avaient des valeurs de l'indice d'éloignement de 0,10 à 0,14. Toutes ces SDR se trouvaient au Québec, en Ontario, en Alberta et en Colombie-Britannique, à proximité de régions métropolitaines densément peuplées. Bien qu'elles aient une population de taille faible à moyenne, ces SDR avaient de faibles valeurs de l'indice d'éloignement en raison de leur proximité avec de grandes RMR.

2. Le coût de déplacement était fondé sur le coût du déplacement d'une personne entre deux endroits, selon l'option de déplacement la plus accessible et la moins coûteuse.

Figure 1
Répartition des valeurs de l'indice d'éloignement des subdivisions de recensement au Canada



Source : Les données de ce graphique sont tirées du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017).

Lors des travaux préliminaires, on a étudié le nombre optimal de catégories pour chacune des méthodes de classification. On a classé les valeurs de l'indice d'éloignement continu dans différentes catégories à l'aide de diverses méthodes. Pour chacune de ces méthodes, on a obtenu entre 3 et 10 catégories distinctes. Lorsque les valeurs de l'indice d'éloignement étaient classées en moins de cinq catégories, il était difficile de distinguer les régions urbaines, rurales et éloignées précisément en raison de la géographie du Canada, qui va d'importants centres urbains à des régions très éloignées ne comptant aucun habitant. De plus, les approches qui comptaient moins de cinq catégories se rapprochaient des approches traditionnelles de la CSS et des CTRPOP pour ce qui est du nombre de catégories.

En revanche, lorsque plus de cinq catégories étaient formées, cela faisait en sorte que la répartition des SDR et des populations était disproportionnée, certaines catégories comptant seulement quelques SDR et de petites populations. Ces grappes comptant de petites populations et quelques SDR ne se prêtent pas au calcul de l'indice d'éloignement, puisqu'elles ne permettent pas de comparer les SDR de différentes catégories d'éloignement. Par exemple, dans l'analyse par grappes à K-moyennes, les valeurs de l'indice d'éloignement ont été divisées en 10 catégories, ce qui fait que certaines des catégories ont été formées au moyen de quelques SDR ayant une petite population.

De plus, la formation de plus de cinq catégories faisait croître la probabilité que des renseignements privés et confidentiels soient divulgués, surtout lorsque venait le temps de diffuser des données provenant d'enquêtes auprès d'un échantillon de petite taille ou de déclarer des données sur des conditions rares. Par conséquent, il a été décidé que les analyses principales porteraient uniquement sur les classifications comportant cinq catégories, à savoir les « régions facilement accessibles », les « régions accessibles », les « régions moins accessibles », les « régions éloignées » et les « régions très éloignées ». On a utilisé le logiciel SAS Enterprise Guide 7.1 pour classer les données, et le logiciel ArcMap 10.5.1 pour illustrer la répartition géographique des SDR par catégorie dans les cartes.

3 Résultats

L'éloignement est un concept relatif et non absolu. Le principal défi dans le processus de classification consiste à déterminer les points de démarcation (seuils) entre les catégories. Comme il n'y a pas de critères établis pour classer une variable continue en catégories distinctes en fonction de l'éloignement, les approches de classification proposées dans le présent document ont été choisies en fonction d'une revue approfondie de la littérature. Cinq approches de classification, chacune comportant cinq catégories (régions facilement accessibles, régions accessibles, régions moins accessibles, régions éloignées et régions très éloignées), ont été examinées et comparées en fonction de leur spécificité, de leur objectif de recherche et de leur pertinence pour la diffusion des données.

3.1 Classification manuelle des valeurs de l'indice d'éloignement en fonction des seuils naturels, des chiffres de population et de la répartition des subdivisions de recensement

La méthode de classification manuelle divise manuellement les valeurs de l'indice d'éloignement en cinq catégories mutuellement exclusives, tout en tenant compte des seuils naturels observés dans la répartition des valeurs de l'indice d'éloignement (figure 1). La classification manuelle examine la répartition des valeurs de l'indice d'éloignement et repère manuellement les seuils naturels dans les données, tout en tenant compte du nombre de SDR et de la répartition de la population dans chaque catégorie. Près de 88 % des SDR avaient des valeurs de l'indice d'éloignement inférieures à 0,55 (sur une échelle de 0 à 1 où 0 représente les régions les plus accessibles et 1 représente les régions les moins accessibles). Par conséquent, quatre des cinq catégories ont été formées entre les valeurs de l'indice d'éloignement de 0 et de 0,55. La seule catégorie qui dépassait le seuil de 0,55 était celle des « régions très éloignées ». Toutes les SDR des catégories « régions facilement accessibles » et « régions accessibles » étaient rattachées à un réseau routier ou de traversiers.

Tableau 1

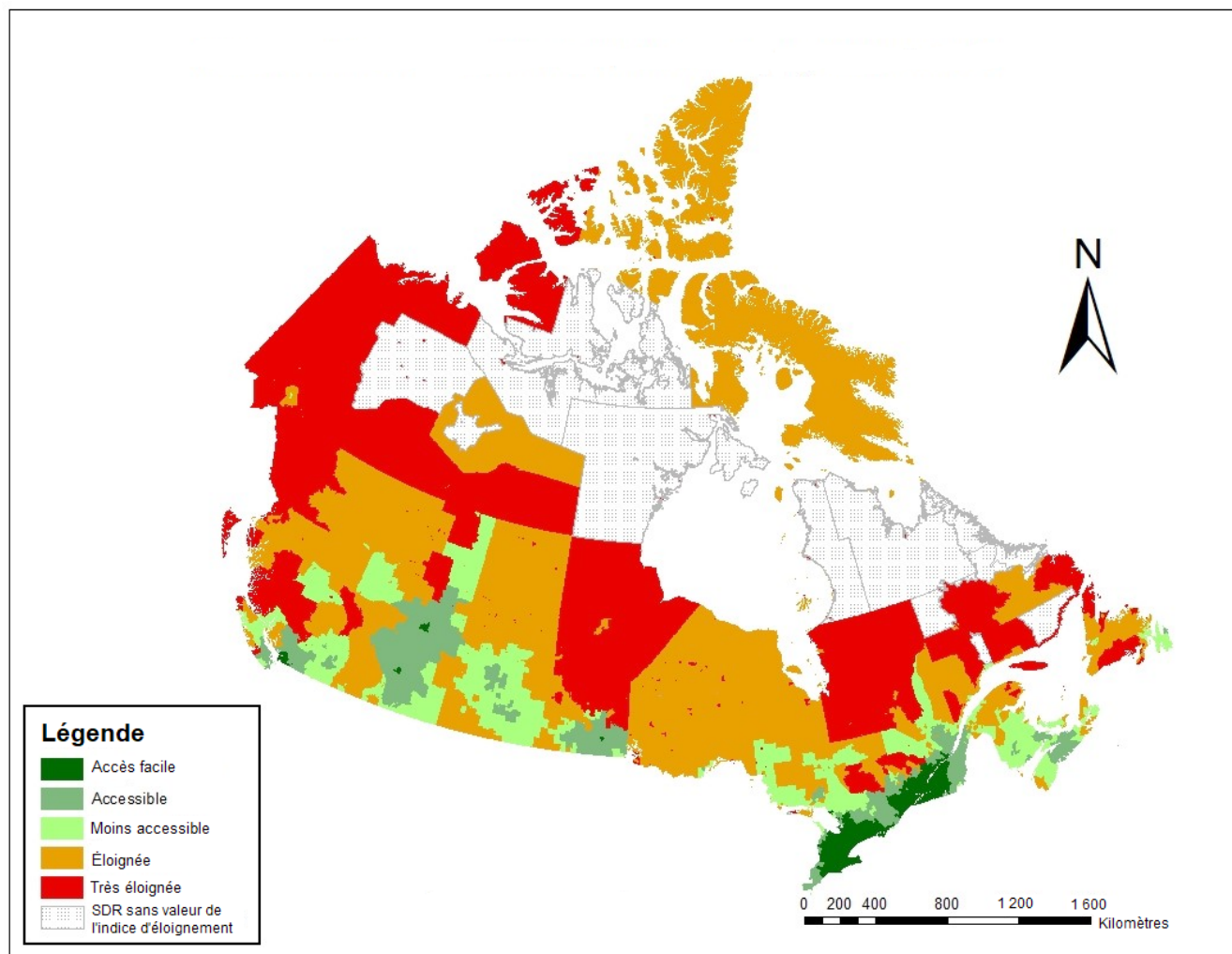
Classification manuelle de l'indice d'éloignement en fonction des seuils naturels, de la population et du nombre de subdivisions de recensement

Catégorie d'éloignement	Valeur de l'indice d'éloignement	Population	SDR	Population
		pourcentage		moyenne par SDR
Régions facilement accessibles	< 0,1500	68,14	15,10	30 974
Régions accessibles	De 0,1500 à 0,2888	19,34	21,48	6 180
Régions moins accessibles	De 0,2889 à 0,3898	7,90	27,51	1 972
Régions éloignées	De 0,3899 à 0,5532	3,84	23,98	1 098
Régions très éloignées	> 0,5532	0,78	11,92	449

Notes : SDR signifie « subdivision de recensement ». Les seuils naturels sont utilisés de telle sorte que toutes les SDR comptant plus de 100 000 habitants soient classées dans la catégorie « régions facilement accessibles » ou « régions accessibles », et que toutes les SDR comptant plus de 20 000 habitants soient classées dans la catégorie « régions facilement accessibles », « régions accessibles » ou « régions moins accessibles ». Les SDR comptant plus de 10 000 habitants ne sont pas classées dans la catégorie « régions très éloignées ».

Source : Les données de ce tableau sont tirées du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017).

La population totale dans chaque catégorie a également été prise en compte dans la classification des valeurs de l'indice d'éloignement, malgré le fait qu'environ 82 % de la population canadienne vit dans des régions urbaines (Statistique Canada, 2016c). Les seuils ont été établis manuellement de sorte que toutes les SDR dont la population est supérieure à 100 000 habitants soient classées dans les catégories « régions facilement accessibles » ou « régions accessibles ». Cette classification concordait avec la méthode de la CSS, selon laquelle les CTRPOP dont la population est supérieure à 100 000 habitants (et dont le noyau compte au moins 50 000 habitants) sont considérés comme des RMR. De même, toutes les SDR ayant une population de plus de 20 000 habitants ont été classées dans les catégories « régions facilement accessibles », « régions accessibles » ou « régions moins accessibles ». Les SDR de plus de 10 000 habitants n'ont pas été classées dans la catégorie « régions très éloignées ».

Carte 1**Classification manuelle de l'indice d'éloignement en fonction des seuils naturels, de la population et du nombre de subdivisions de recensement dans chaque catégorie**

Note : SDR signifie « subdivision de recensement ».

Source : Les auteurs de la présente étude ont élaboré la carte 1 à l'aide du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017) et du Fichier des limites des subdivisions de recensement de Statistique Canada (lcsd000b16a_f.shp).

L'un des principaux avantages de cette méthode de classification est qu'elle sépare les données en catégories fondées sur des groupes naturels dans la répartition des données, sans compromettre la répartition du nombre de SDR dans chaque catégorie. Par exemple, elle classe au moins 12 % des SDR dans chaque catégorie. Cependant, cette méthode de classification est très subjective. Les seuils sont sélectionnés manuellement et, par conséquent, doivent être mis à jour manuellement en fonction des nouvelles régions géographiques et des nouveaux renseignements sur la population chaque fois qu'un recensement est mené. Un autre problème lié à cette méthode est que la répartition de la population selon qu'elle vit en région urbaine ou rurale est différente de ce que l'on sait déjà sur la répartition de la population au Canada. Les catégories « régions facilement accessibles » et « accessibles » comptent à elles seules plus de 87 % de la population totale, ce qui est supérieur à la population de l'ensemble des CTRPOP observée au Canada (Statistique Canada, 2016c). Cela dit, comme la classification des valeurs de l'indice d'éloignement est fondée sur une approche différente de celle de l'approche des CTRPOP, la répartition de la population peut différer selon la catégorie. Les catégories d'éloignement du tableau 1 ont été élaborées en fonction des critères susmentionnés.

3.2 Classification par intervalles égaux fondée sur la valeur de l'indice d'éloignement

Une autre méthode de classification des valeurs de l'indice d'éloignement est la méthode de classification par intervalles égaux, dans laquelle on crée cinq catégories d'éloignement en divisant les valeurs de l'indice d'éloignement de façon égale en quintiles. Cette méthode de classification divise les valeurs des attributs — dans ce cas, les valeurs de l'indice d'éloignement — en intervalles de valeurs égales. Cette méthode est simple et souple, car le nombre de catégories peut varier selon les besoins de l'utilisateur. Cette méthode convient également à l'affichage de données qui varient linéairement, comme les valeurs de l'indice d'éloignement, et qui ne comptent pas de vraies valeurs aberrantes.

Tableau 2
Classification par intervalles égaux selon la valeur de l'indice d'éloignement

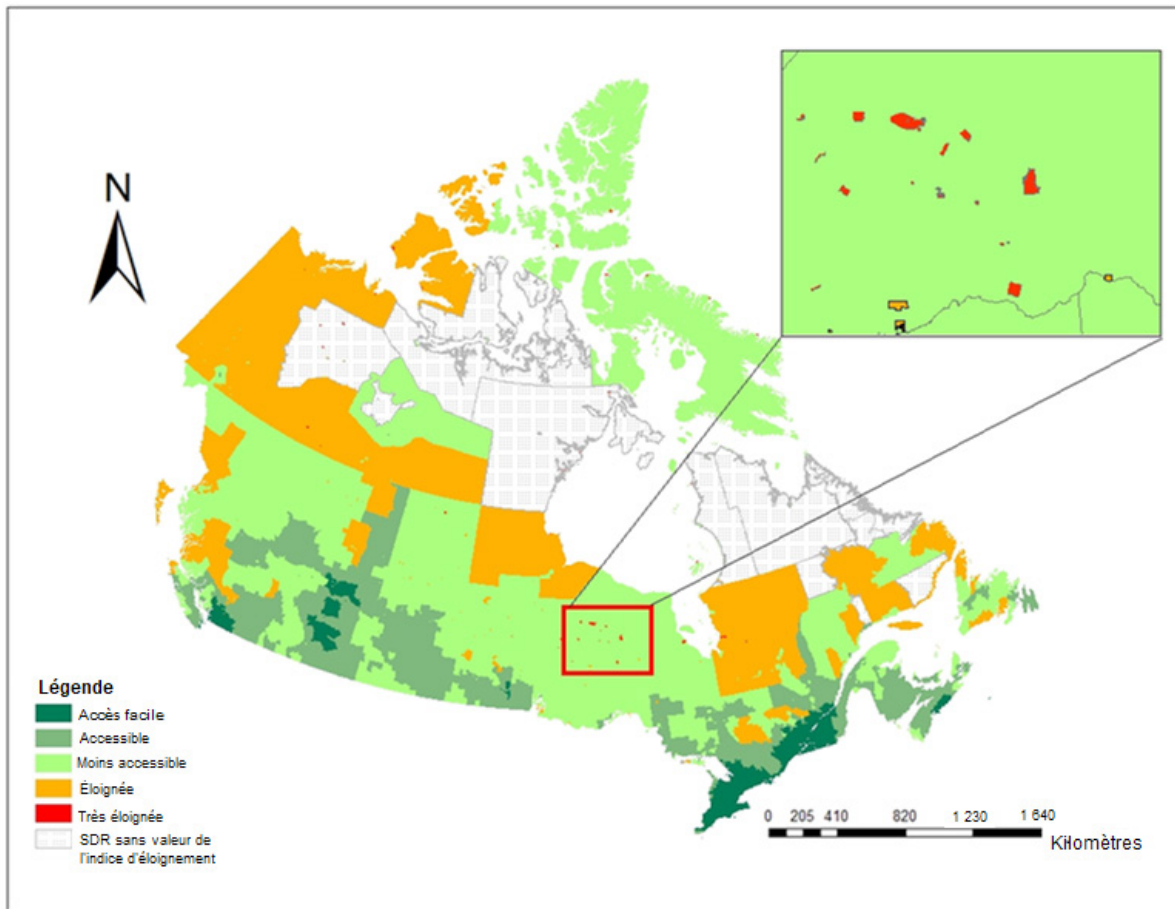
Catégorie d'éloignement	Valeur de l'indice d'éloignement	Population	SDR	Population
		pourcentage		moyenne par SDR
Régions facilement accessibles	< 0,20	75,36	21,01	24 618
Régions accessibles	De 0,20 à 0,3999	20,56	45,66	3 091
Régions moins accessibles	De 0,40 à 0,5999	3,59	24,80	992
Régions éloignées	De 0,60 à 0,7999	0,35	6,99	344
Régions très éloignées	> ou = 0,80	0,14	1,54	642

Notes : SDR signifie « subdivision de recensement ». Les valeurs de l'indice d'éloignement des SDR sont réparties également en cinq catégories (classification des valeurs de l'indice d'éloignement par quintile).

Source : Les données de ce tableau sont tirées du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017).

Toutefois, cette méthode est moins appropriée que la méthode de classification manuelle parce que la classification est fondée sur la valeur de l'indice d'éloignement et non sur les seuils naturels dans les données. De plus, cette méthode ne tient pas compte de la population et du nombre de SDR dans chaque catégorie. Par exemple, plus de 96 % de la population a été classée dans les catégories « régions facilement accessibles » ou « accessibles », ce qui ne concorde pas avec les autres classifications des CTRPOP utilisées au Canada. En général, la population moyenne par SDR diminue à mesure que l'éloignement augmente. Toutefois, selon cette classification, la population moyenne par SDR dans la catégorie « régions éloignées » (344) était inférieure à la population moyenne par SDR dans la catégorie « régions très éloignées » (642) parce qu'il y avait peu de SDR dans la catégorie « régions très éloignées ».

Carte 2
Classification par intervalles égaux selon la valeur de l'indice d'éloignement



Note : SDR signifie « subdivision de recensement ».

Source : Les auteurs de la présente étude ont élaboré la carte 2 à l'aide du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017) et du Fichier des limites des subdivisions de recensement de Statistique Canada (lcsd000b16a_f.shp).

En revanche, plus de 45 % des SDR étaient classées dans la catégorie « régions accessibles », tandis que moins de 2 % étaient classées dans la catégorie « régions très éloignées » (certaines de ces RMR sont illustrées dans la section agrandie de la carte 2). Par conséquent, certaines collectivités très éloignées du Nord du Canada n'ont pas été classées comme étant « très éloignées » parce que la valeur de l'indice d'éloignement ne dépassait pas le seuil de 0,8.

3.3 Classification par quantiles en fonction du nombre de subdivisions de recensement

Une autre méthode de classification des valeurs de l'indice d'éloignement est la classification par quantiles en fonction du nombre total de SDR. Dans cette méthode, les SDR sont réparties en nombre égal dans chacune des cinq catégories. Un avantage important de cette méthode est qu'elle répartit les SDR de façon égale dans chaque catégorie d'éloignement, et chaque catégorie est représentée de façon égale sur la carte. Cette méthode est plus souple que d'autres puisqu'elle permet de créer plus de cinq ou moins de cinq catégories en divisant simplement le nombre de SDR en parts égales.

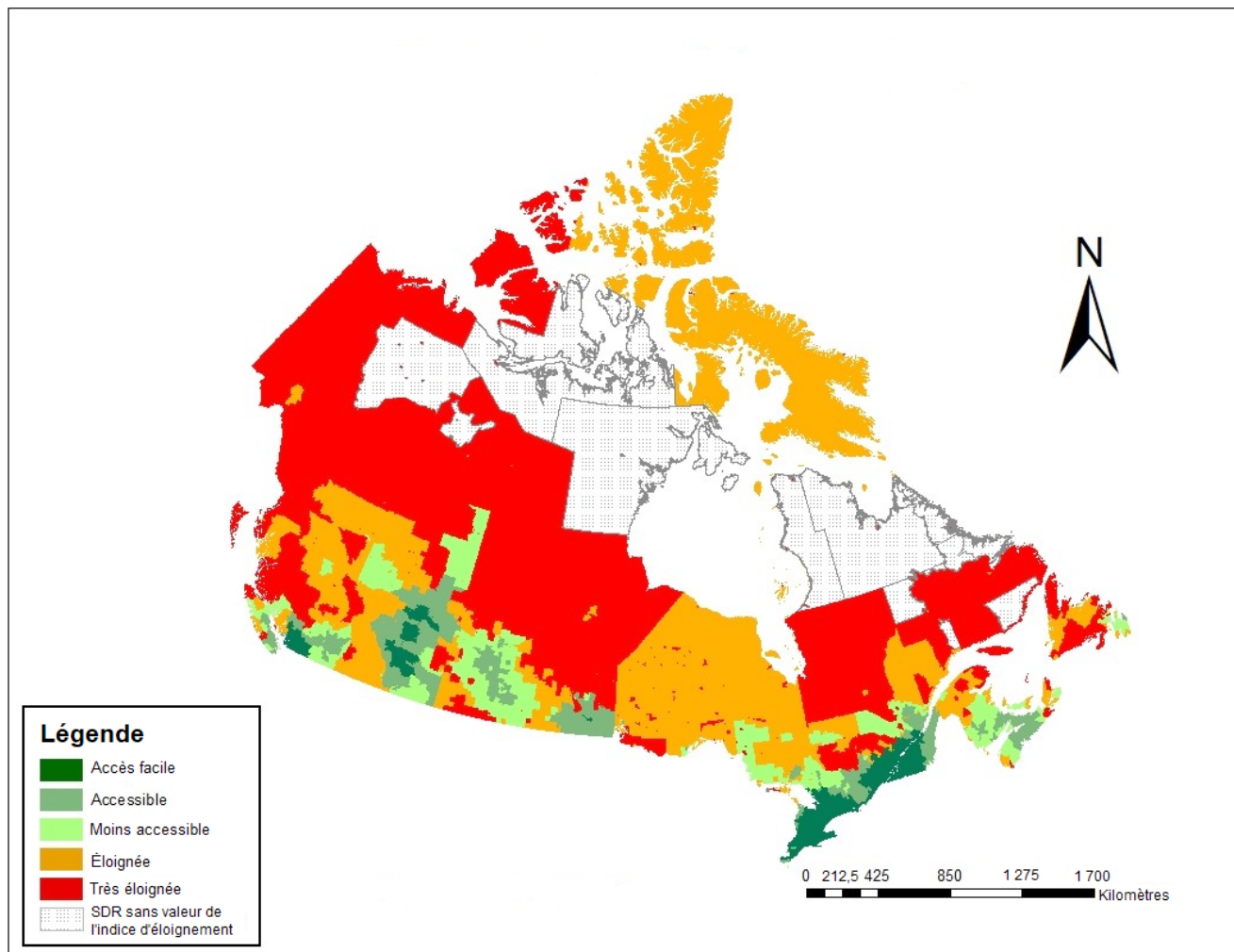
Tableau 3
Classification par quantiles en fonction du nombre de subdivisions de recensement

Catégorie d'éloignement	Valeur de l'indice d'éloignement	Population	SDR	Population
		pourcentage		moyenne par SDR
Régions facilement accessibles	< 0,1928	73,43	20,00	25 204
Régions accessibles	De 0,1928 à 0,3013	15,50	20,00	5 319
Régions moins accessibles	De 0,3014 à 0,3739	5,11	20,00	1 753
Régions éloignées	De 0,3740 à 0,4744	4,16	20,00	1 432
Régions très éloignées	> 0,4744	1,80	20,00	617

Notes : SDR signifie « subdivision de recensement ». Le nombre de SDR est réparti également en cinq catégories (classification par quantiles selon la SDR).

Source : Les données de ce tableau sont tirées du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017).

Toutefois, la carte ou le tableau de données qui en résulte peut être trompeur parce que des SDR ayant des valeurs de l'indice d'éloignement semblables peuvent se retrouver dans différentes catégories, et des SDR d'une même catégorie peuvent avoir des valeurs de l'indice d'éloignement différentes. Les SDR éloignées sont généralement plus grandes que les SDR urbaines. Par conséquent, la carte est trompeuse parce que, avec cette méthode, la catégorie « régions très éloignées » est surreprésentée. Cette méthode de classification compromet également la répartition des valeurs de l'indice d'éloignement et de la population dans chaque catégorie. Comme elle est fondée sur le nombre de SDR, cette méthode est moins utile pour les études fondées sur la population. Un autre inconvénient de cette méthode est que de nombreuses SDR ayant une population de plus de 10 000 habitants sont classées dans la catégorie « régions très éloignées » puisque la classification est fondée sur le nombre de SDR et non sur la valeur réelle de l'indice d'éloignement. Par conséquent, cette classification pourrait ne pas représenter fidèlement la population réelle des régions éloignées du Canada.

Carte 3**Classification par quantiles en fonction du nombre de subdivisions de recensement**

Note : SDR signifie « subdivision de recensement ».

Source : Les auteurs de la présente étude ont élaboré la carte 3 à l'aide du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017) et du Fichier des limites des subdivisions de recensement de Statistique Canada (lcsd000b16a_f.shp).

3.4 Classification en fonction des seuils naturels de Jenks

La méthode de classification en fonction des seuils naturels de Jenks est fondée sur l'algorithme des seuils naturels de Jenks (Jenks, 1967; Brewer et Pickle, 2002). Cette méthode crée des catégories fondées sur les seuils naturels dans les données en regroupant les valeurs semblables et en maximisant les écarts entre les catégories.

Plus précisément, cette méthode de classification vise à réduire au minimum l'écart moyen de chacune des valeurs au sein d'une catégorie par rapport à la moyenne des valeurs de la catégorie, tout en maximisant l'écart moyen de chaque catégorie par rapport aux moyennes des autres catégories (Brewer et Pickle, 2002). Par conséquent, les variances au sein de chaque catégorie sont réduites au minimum, tandis qu'elles sont maximisées entre les catégories.

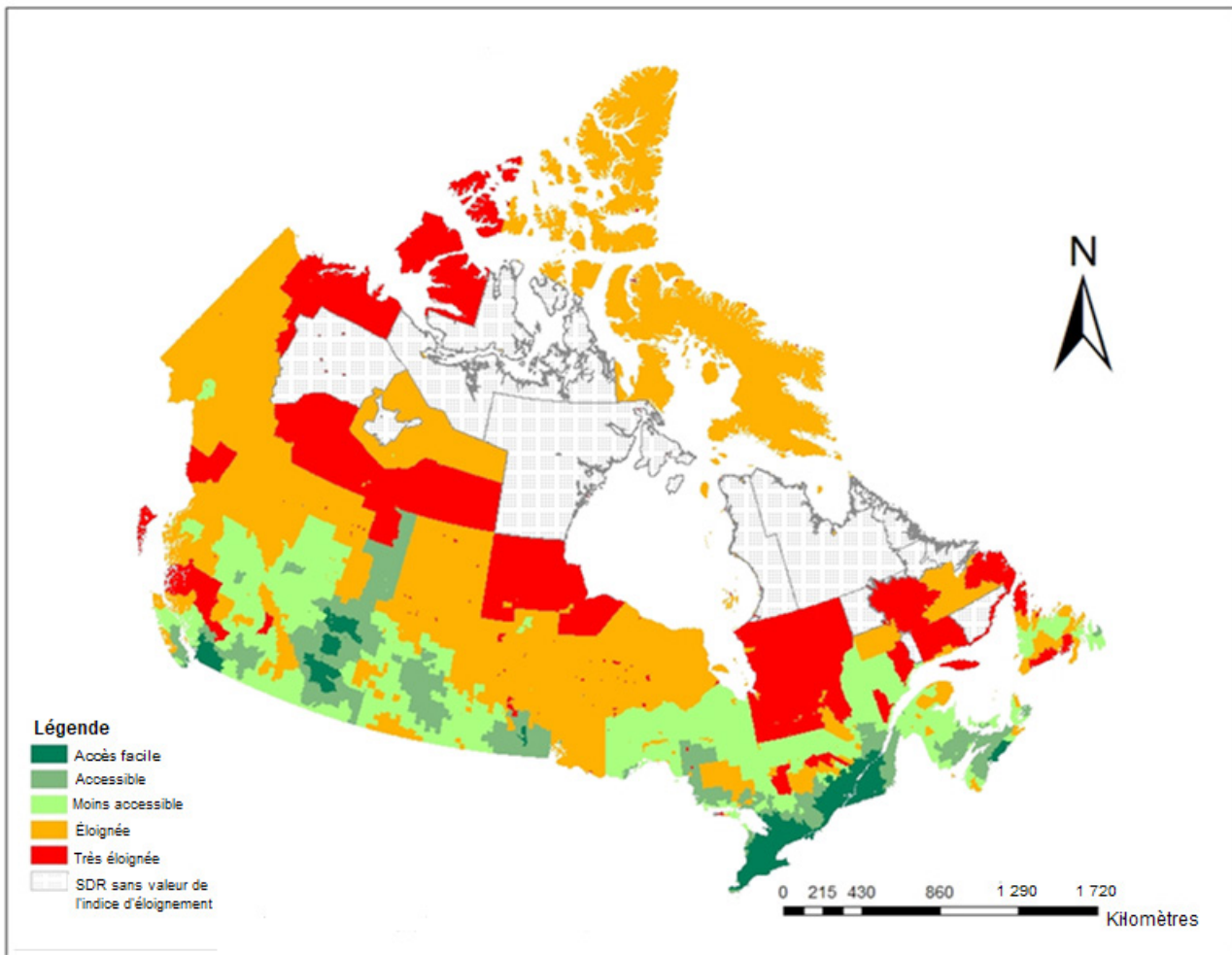
Tableau 4
Classification de l'indice d'éloignement en fonction des seuils naturels de Jenks

Catégorie d'éloignement	Valeur de l'indice d'éloignement	Population	SDR	Population moyenne par SDR
		pourcentage		
Régions facilement accessibles	< 0,2002	75,37	21,03	24 596
Régions accessibles	De 0,2002 à 0,3269	16,14	25,13	4 408
Régions moins accessibles	De 0,3270 à 0,4546	6,27	31,20	1 379
Régions éloignées	De 0,4547 à 0,6267	1,80	15,40	802
Régions très éloignées	> 0,6267	0,43	7,24	411

Notes : SDR signifie « subdivision de recensement ». L'algorithme des seuils naturels de Jenks est appliqué. Les catégories sont formées en fonction des seuils naturels dans les données.

Source : Les données de ce tableau sont tirées du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017).

Carte 4
Classification de l'indice d'éloignement en fonction des seuils naturels de Jenks



Note : SDR signifie « subdivision de recensement ».

Source : Les auteurs de la présente étude ont élaboré la carte 4 à l'aide du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017) et du Fichier des limites des subdivisions de recensement de Statistique Canada (lcsd000b16a_f.shp).

Cette méthode classe environ 91 % de la population dans les catégories « régions facilement accessibles » et « régions accessibles ». Cette part est supérieure à l'estimation de la population des CTRPOP (Statistique Canada, 2016c), mais il ne faut pas s'en étonner, puisque l'indice d'éloignement constitue une méthode différente visant à comprendre les populations urbaines et rurales au Canada. Bien que le nombre de SDR et la population ne soient pas répartis proportionnellement dans chaque catégorie, cette méthode est davantage scientifique parce que les catégories sont formées en fonction des caractéristiques des données héritées. Cette méthode permet en outre de mettre facilement à jour les données lorsque celles-ci sont modifiées.

3.5 Classification des valeurs de l'indice d'éloignement par K-moyennes

La dernière méthode de classification étudiée était une analyse des valeurs de l'indice d'éloignement par grappes à l'aide d'une méthode de classification par K-moyennes (Macqueen, 1967; Lloyd, 1982). Il s'agit d'une méthode de regroupement des données conçue pour déterminer la meilleure classification des valeurs en « k » catégories (grappes). L'algorithme de classification par K-moyennes estime de façon itérative les moyennes des grappes et attribue chaque observation à la grappe ayant la moyenne la plus proche (Tan, Steinbach et Kumar, 2019). Autrement dit, cette méthode vise à réduire la variance au sein des catégories et à maximiser la variance entre les catégories.

Différents nombres de grappes, d'itérations et de méthodes de remplacement des valeurs initiales sont disponibles dans l'algorithme de classification par K-moyennes. Par conséquent, pour trouver le modèle le plus approprié pour notre ensemble de données, on a appliqué différentes méthodes de classification par K-moyennes pour créer de 3 à 10 grappes en effectuant de 1 à 5 itérations en utilisant des méthodes de remplacement intégral et aléatoire des valeurs initiales des moyennes de grappe. Le modèle le plus approprié a produit cinq grappes issues de cinq itérations et de l'application d'une méthode de remplacement intégral des valeurs initiales des moyennes de grappe (tableau 5).

Tableau 5
Classification de l'indice d'éloignement par K-moyennes

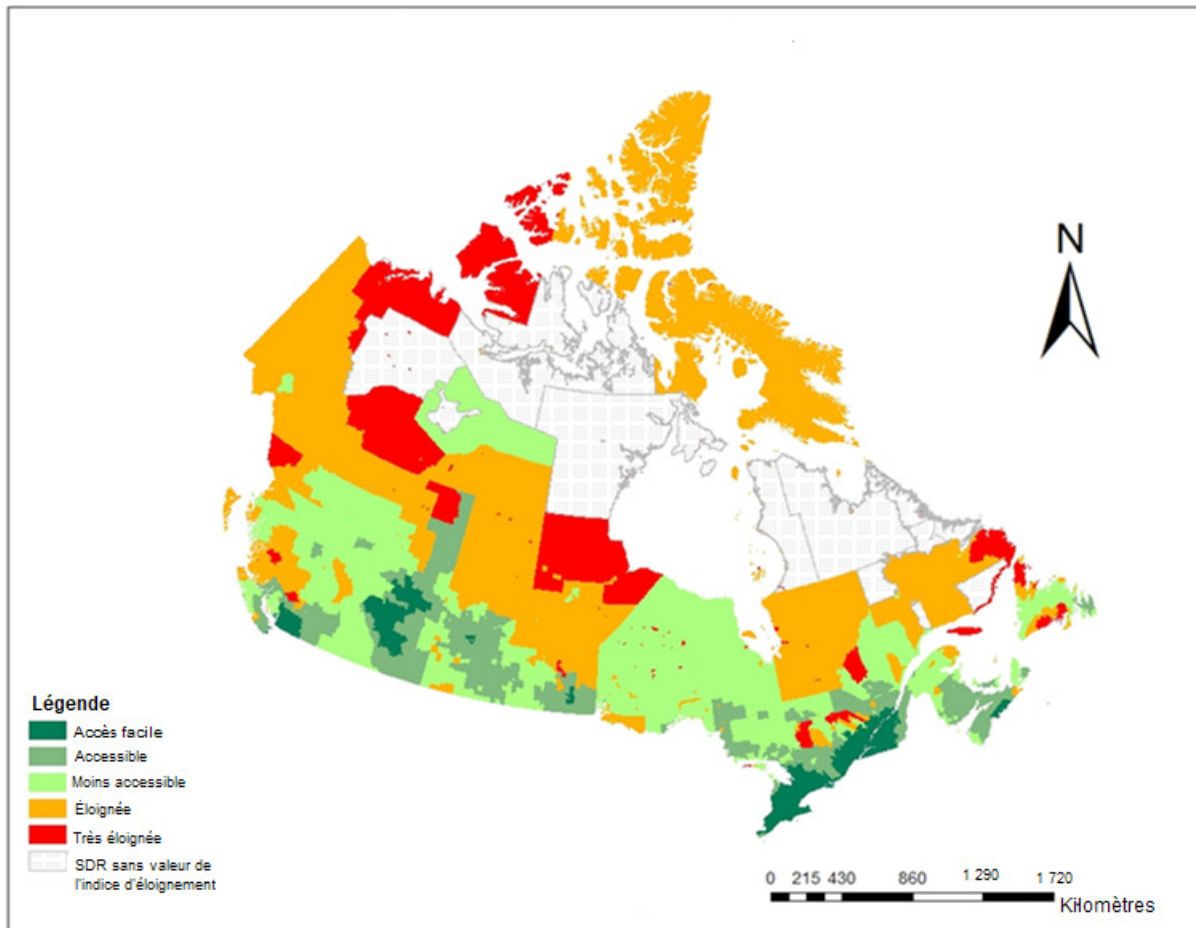
Catégorie d'éloignement	Valeur de l'indice d'éloignement	Population	SDR	Population
		pourcentage		moyenne par SDR
Régions facilement accessibles	< 0,2130	76,59	22,93	22 931
Régions accessibles	De 0,2131 à 0,3565	16,78	32,62	3 531
Régions moins accessibles	De 0,3566 à 0,5080	5,31	28,33	1 288
Régions éloignées	De 0,5081 à 0,6877	1,01	11,47	602
Régions très éloignées	> 0,6877	0,31	4,65	461

Notes : SDR signifie « subdivision de recensement ». On utilise une méthode d'analyse par grappes à K-moyennes pour former de 3 à 10 grappes en effectuant de 1 à 5 itérations avec remplacement intégral et aléatoire des valeurs initiales. Le modèle le plus approprié est celui qui comporte cinq grappes et cinq itérations avec remplacement intégral des valeurs initiales.

Source : Les données de ce tableau sont tirées du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017).

Cette méthode est raisonnablement efficace puisqu'elle tient compte des seuils naturels dans les données et classe chaque SDR dans le groupe le plus proche. Il s'agit de l'une des méthodes les plus souples, car l'utilisateur peut choisir différentes options de regroupement de même que le modèle qui convient le mieux à ses ensembles de données. Cette méthode de classification se prête mieux que d'autres à la classification des valeurs de l'indice d'éloignement parce qu'elle permet le traitement de données ordinales. Cette méthode utilise un algorithme différent et présente un avantage par rapport à la méthode des seuils naturels de Jenks, car elle permet de traiter plusieurs variables en même temps pour créer des grappes. La méthode de classification par K-moyennes est particulièrement utile lorsque les données sont ventilées en de nombreux groupes.

Carte 5 Classification de l'indice d'éloignement par K-moyennes



Note : SDR signifie « subdivision de recensement ».

Source : Les auteurs de la présente étude ont élaboré la carte 5 à l'aide du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017) et du Fichier des limites des subdivisions de recensement de Statistique Canada (lcsd000b16a_f.shp).

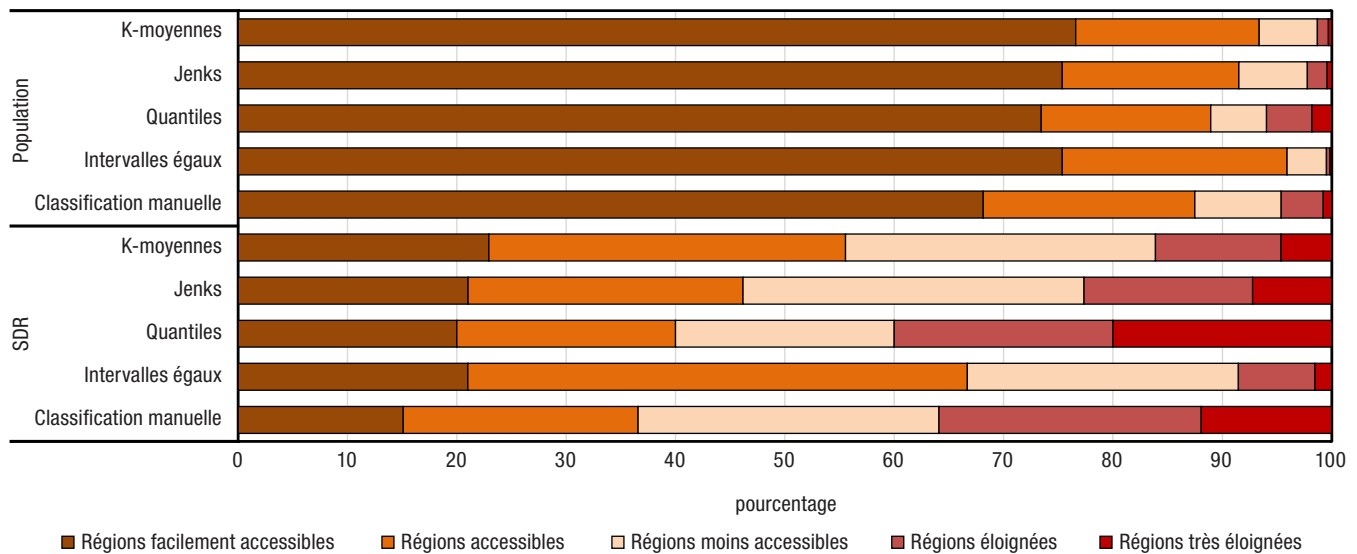
L'un des principaux inconvénients de cette méthode est qu'elle est sensible à l'échelle. Cependant, les valeurs des indices d'éloignement sont déjà normalisées de 0 à 1, de sorte qu'il est facile de les classer par K-moyennes. Pour l'ensemble de données de l'indice d'éloignement, une seule variable (valeurs de l'indice d'éloignement) est utilisée pour l'analyse par grappes parce que l'inclusion d'autres variables (p. ex. la taille de la population) entrerait en conflit avec l'échelle ordinaire de 0 à 1 de l'indice d'éloignement. De plus, comme la taille de la population est déjà utilisée pour calculer les valeurs de l'indice d'éloignement, il serait redondant de l'utiliser de nouveau dans la classification par K-moyennes.

4 Conclusions et recommandations

Bien qu'il n'existe pas de classification des régions selon qu'elles sont urbaines ou rurales qui soit universellement acceptée, les pays ont tenté d'établir leur propre classification en fonction de leurs caractéristiques géographiques, environnementales, démographiques et sociopolitiques. Comme le Canada est un grand pays comportant des régions éloignées, la classification des régions selon qu'elles sont urbaines ou rurales revêt une importance dans la collecte de renseignements sur les besoins de chaque collectivité en matière de biens, de services et de soins de santé. L'indice d'éloignement élaboré par Alasia et coll. (2017) fournit une valeur approximative raisonnable de l'éloignement relatif des collectivités canadiennes. L'indice d'éloignement saisit également une dimension géographique unique, à savoir l'accessibilité, qui se distingue des dimensions saisies par d'autres classifications géographiques, comme la CSS et la classification des CTRPOP.

Graphique 1

Comparaison des méthodes de classification de l'indice d'éloignement selon la population et le nombre de subdivisions de recensement



Note : SDR signifie « subdivision de recensement ».

Source : Les données de ce graphique sont calculées à partir du tableau de données de l'indice d'éloignement des subdivisions de recensement canadiennes (Alasia et coll., 2017).

Toutefois, les applications de recherche en ce qui concerne l'indice d'éloignement sont limitées en raison d'un manque de catégorisation. Le présent document aborde cinq approches méthodologiques différentes pour classer les valeurs de l'indice d'éloignement en catégories distinctes afin que les utilisateurs puissent choisir une méthode de classification qui répond à leurs besoins. Les méthodes de classification ont été choisies à la suite d'une revue de la littérature sur les pratiques exemplaires, et orientées par les méthodes de classification utilisées dans d'autres pays.

Parmi les cinq méthodes de classification présentées dans ce document, la méthode de classification manuelle, la méthode de classification en fonction des seuils naturels de Jenks et la méthode de classification par K-moyennes sont plus appropriées et plus fiables que la méthode de classification par quantiles et la méthode de classification par intervalles égaux pour ce qui est de la répartition de la population et des SDR dans chaque catégorie. Comme le montre le graphique 1, les méthodes de classification par quantiles et par intervalles égaux représentent les deux extrémités du spectre. Le nombre de SDR et la population totale dans la catégorie « régions très éloignées » ont été surestimés par la méthode de classification par quantiles et sous-estimés par la méthode de classification par intervalles égaux.

Les résultats peuvent varier selon la méthode de classification utilisée, puisque les seuils et la répartition de la population dans les régions urbaines et rurales diffèrent d'une méthode à l'autre. Toutefois, la méthode de classification manuelle, la méthode de classification en fonction des seuils naturels de Jenks et la méthode de classification par K-moyennes offrent plus d'avantages que les autres méthodes de classification étudiées dans le présent document. La méthode manuelle est souple et tient compte de la proportion de la population, de la répartition des SDR et de la valeur de l'indice d'éloignement. Les méthodes de classification en fonction des seuils naturels de Jenks et par K-moyennes visent à réduire au minimum l'écart moyen de chaque valeur au sein d'une catégorie par rapport à la moyenne des valeurs de la catégorie, tout en maximisant l'écart moyen de chaque catégorie par rapport aux moyennes des autres catégories. Ces méthodes se prêtent mieux à la classification des régions géographiques du Canada. Par conséquent, les auteurs de l'étude recommandent ces trois méthodes de classification aux fins d'utilisation future. Soulignons que des chercheurs ont déjà mis la méthode de classification manuelle à profit pour étudier la variabilité géographique des taux de mortalité évitable au Canada (Subedi, Greenberg et Roshanafshar, 2019). Dans l'avenir, les chercheurs devront tenir compte des changements dans l'éloignement relatif d'une région et des changements possibles dans les données d'entrée de l'indice d'éloignement (limites géographiques et populations des SDR) au moment de déterminer la méthode à utiliser.

Bibliographie

- Alasia, A., F. Bédard, J. Bélanger, E. Guimond et C. Penney. 2017. [*Mesurer l'éloignement et l'accessibilité : Un ensemble d'indices applicables aux collectivités canadiennes*](#). Rapports sur les projets spéciaux sur les entreprises. Produit n° 18-001-X au catalogue de Statistique Canada. Ottawa : Statistique Canada. Disponible au lien suivant : <https://www150.statcan.gc.ca/n1/pub/18-001-x/18-001-x2017002-fra.htm>.
- Bibby, P. et P. Brindley. 2013. [*Urban and Rural Area Definitions for Policy Purposes in England and Wales: Methodology*](#). Service statistique du gouvernement. Documents méthodologiques. Londres : Bureau de la statistique nationale. Disponible au lien suivant : https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/239477/RUC11methodologypaperaug_28_Aug.pdf.
- Brewer, C.A., et L. Pickle. 2002. « Evaluation of methods for classifying epidemiological data on choropleth maps in series ». *Annals of the Association of American Geographers* 92(4) : 662 à 681.
- Desmeules, M. et R. Pong. 2006. *Comment se portent les Canadiens vivant en milieu rural? Une évaluation de leur état de santé et des déterminants de la santé*. Initiative sur la santé de la population canadienne. Ottawa : Institut canadien d'information sur la santé.
- Du Plessis, V., R. Beshiri, R.D. Bollman et H. Clemenson. 2002. *Définitions de « rural »*. Série de documents de travail sur l'agriculture et le milieu rural, n° 61. Ottawa : Statistique Canada.
- Eckert, K.A., A.W. Taylor et D. Wilkinson. « Does health service utilization vary by remoteness? South Australian population data and the Accessibility and Remoteness Index of Australia ». *Australian and New Zealand Journal of Public Health* 28 (5) : 426 à 432.
- Hart, G.L., D.M. Lishner et E.H. Larson. 2005. « Rural definitions for health policy and research ». *American Journal of Public Health* 95 (7) : 1149 à 1155.
- Institut australien pour la santé et le bien-être. 2004. [*Rural, Regional and Remote Health: A Guide to Remoteness Classifications*](#). Canberra : Institut australien pour la santé et le bien-être. Disponible au lien suivant : <https://www.aihw.gov.au/getmedia/9c84bb1c-3ccb-4144-a6dd-13d00ad0fa2b/rrrh-gtrc.pdf.aspx?inline=true>.
- Institut canadien d'information sur la santé (ICIS). 2012. « Disparités sur le plan de l'expérience en matière de soins de santé primaires vécue par les Canadiens présentant des conditions propices aux soins ambulatoires ». *Analyse en bref*. Ottawa : Institut canadien d'information sur la santé.
- Jenks, G.F. 1967. « The data model concept in statistical mapping ». *International Yearbook of Cartography* 7 : 186 à 190.
- Johnson, J.A., et A.M. Johnson. 2015. « Urban-rural differences in childhood and adolescent obesity in the United States: A systematic review and meta-data analysis ». *Childhood Obesity* 11 (3) : 223 à 241.
- Lloyd, S.P. 1982. « Least squares quantization in PCM ». *IEEE Transactions on Information Theory* 28 (2) : 129 à 137.
- MacQueen, J. 1967. « Some methods for classification and analysis of multivariate observations ». Dans *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, p. 281 à 297. Berkeley, Californie : University of California Press.
- Ministère de la Santé et des Soins aux aînés. 2001. *Measuring Remoteness: Accessibility/Remoteness Index of Australia (ARIA)*. Occasional Papers New Series, n° 14. Canberra : gouvernement australien, ministère de la Santé.
- Ministère des Industries primaires et de l'Énergie (MIPE) et ministère des Services humains et de la Santé (MSHS). 1994. *Rural, Remote and Metropolitan Areas Classification 1991 Census Edition*. Canberra : Australian Government Publishing Service, p. 1 à 28.
- Mitura, V. et R.D. Bollman. 2003. « The health of rural Canadians: A rural-urban comparison of health indicators ». *Rural and Small Town Canada Analysis Bulletin* 4 (6) : 1 à 23.
- National Center for Health Statistics (NCHS). 2014. *2013 NCHS Urban-Rural Classification Scheme for Counties*. Hyattsville, Maryland : National Center for Health Statistics

Peen, J., R.A. Schoevers, A.T. Beekman et J. Dekker. 2010. « The current status of urban-rural differences in psychiatric disorders ». *Acta Psychiatrica Scandinavica* 121 (2) : 84 à 93.

Productivity Commission. 1994. *Rural, Remote and Metropolitan Areas Classification 1991 Census Edition*. Canberra : Australian Government Publishing Service.

Statistique Canada. 2016a. « Centre de population (CTRPOP) », [Dictionnaire, Recensement de la population, 2016](#). Disponible au lien suivant : <https://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo049a-fra.cfm> (consulté le 21 août 2018).

Statistique Canada. 2016b. [Classification des secteurs statistiques – Variante de CGT 2016](#). Ottawa : Statistique Canada. Disponible au lien suivant : https://www23.statcan.gc.ca/imdb/p3VD_f.pl?Function=getVD&TVD=314312 (consulté le 9 juillet 2018).

Statistique Canada. 2016c. [Classification des centres de population et des régions rurales 2016](#). Ottawa : Statistique Canada. Disponible au lien suivant : <https://www.statcan.gc.ca/fra/sujets/norme/ccpr/2016/introduction>.

Statistique Canada. 2017. [Fichier de conversion des codes postaux MO plus \(FCCP+\)](#). Ottawa : Statistique Canada. Disponible au lien suivant : <https://www150.statcan.gc.ca/n1/fr/catalogue/82F0086X2017001>.

Subedi, R., T.L. Greenberg et S. Roshanafshar. 2019. « La géographie a-t-elle une incidence sur la mortalité? Analyse de la mortalité évitable selon l'indice d'éloignement au Canada ». *Rapports sur la santé* 30 (5) : 3 à 15.

Tan, P.-N., M. Steinbach et V. Kumar. 2019. « Cluster analysis: Basic concepts and algorithms ». Dans *Introduction to Data Mining*, sous la direction de P.-N. Tan, M. Steinbach et V. Kumar, p. 525 à 603. New York : Pearson.

Wandl, A.D.I., V. Nadin, W. Zonneveld et R. Rooij. 2014. « Beyond urban-rural classification: Characterising and mapping territories-in-between across Europe ». *Landscape and Urban Planning* 130 (1) : 50 à 63.