Université de Montréal

Défis conceptuels et méthodologiques posés par la définition opérationnelle du milieu : l'exemple des milieux résidentiels favorables à la marche

par

Mylène Riva

Secteur de la santé publique
Faculté de médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiae Doctor (PhD) en Santé publique
option Promotion de la santé

Juillet, 2008

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Défis conceptuels et méthodologiques posés par la définition opérationnelle du milieu : l'exemple des milieux résidentiels favorables à la marche

présentée par :

Mylène Riva

a été évaluée par un jury composé des personnes suivantes :

Marie-France Raynault
présidente-rapporteuse

Lise Gauvin
directrice de recherche

Jean-Marc Brodeur
codirecteur

Philippe Apparicio
codirecteur

Jennifer O'Loughlin
membre du jury

Marius Thériault
examinateur externe

Sylvana Côté
représentante du doyen de la FES

Résumé

Problématique : Malgré un nombre croissant d'études démontrant l'influence significative des milieux de vie sur la santé, définir et opérationnaliser le milieu demeurent problématiques. D'une part, il n'y a pas de consensus sur la définition conceptuelle du milieu. D'autre part, dans une majorité d'études, les milieux sont opérationnalisés par des unités spatiales administratives, tels les secteurs de recensement (SR). L'utilisation de ces unités soulève des enjeux de validité dont celui du degré d'homogénéité des milieux au niveau des expositions. Une absence d'homogénéité des expositions peut mener à des biais de mesure, voire de mauvaise classification (« *misclassification* »), et des biais d'estimation des effets de milieux (« *misestimation* »). Or, jusqu'à présent, la résolution de ce défi méthodologique n'a pas fait l'objet de beaucoup d'études.

Objectif : Cette thèse porte sur la question de recherche suivante : *Quel est l'apport du développement et de l'application d'une approche novatrice pour créer des unités spatiales pour la compréhension des effets de milieux sur la santé ?* Cette approche est fondée sur la création de zones homogènes en termes d'une exposition spécifique, le potentiel de vie active des milieux résidentiels, lequel semble être associé à un indicateur de santé spécifique, la marche.

Méthodes : Les 3206 aires de diffusion de l'île de Montréal caractérisées comme similaires au niveau du potentiel de vie active, mesuré par la densité de population, la mixité d'occupation du sol et l'accessibilité géographique aux services de proximité, ont été regroupées statistiquement en sept catégories optimales d'exposition puis importées dans un système d'information géographique; 898 zones homogènes, caractérisées par une des sept catégories d'exposition, ont ainsi été créées. La valeur de cette approche a été évaluée en quantifiant les degrés d'homogénéité des zones et des SR au niveau du potentiel de vie active. L'influence de ce potentiel sur les habitudes de marche de 2716 adultes a été examinée en comparant les résultats obtenus lorsque le milieu est opérationnalisé par les nouvelles unités spatiales plutôt que par les secteurs de recensement.

Résultats : Alors que les SR sont homogènes en termes d'accessibilité aux services, ils sont hétérogènes en fonction de la densité de la population et de la mixité de l'occupation du sol. De plus, une majorité des SR (55.5%) est caractérisée par une combinaison de trois catégories d'exposition ou plus. La validité des SR pour étudier les déterminants environnementaux de la marche est donc limitée. L'accessibilité géographique aux services de proximité appert comme un déterminant important de la marche tant au niveau des zones que des SR. Bien que la taille de cet effet soit un peu plus grande au niveau des zones, les résultats sont similaires à ceux obtenus en utilisant les SR.

Conclusion : L'approche qui consiste à créer des unités spatiales qui maximisent l'homogénéité des expositions s'avère avantageuse pour définir les milieux, car elle améliore la qualité des inférences et donc la validité interne du devis de recherche. D'autres études sont toutefois requises pour confirmer et mieux documenter l'apport de cette nouvelle approche pour définir et opérationnaliser les milieux de vie.

Mots clés : Effets de milieux, marche, modèles multiniveaux, systèmes d'information géographique.

Summary

Rationale: Although a growing number of studies show significant area effects on health outcomes, defining and operationalising areas remain a challenge. A majority of studies employ administrative spatial unit, such as census tracts to operationalise areas, despite theoretical and empirical issues raised by their use. Moreover, absence of homogeneity of exposures in spatial units of analysis can lead to measurement errors (misclassification) and to biased estimation (misestimation) of area effects. Until now, this important methodological consideration has rarely been addressed.

Objective: The thesis focuses on addressing the following question: *What is the contribution of developing and implementing a novel approach to designing spatial units for understanding area effects on health?* The approach is based on designing areas homogenous in terms of one specific ecologic exposure namely, active living potential, which is hypothesised to be associated with walking.

Methods: The 3206 dissemination areas located on the Island of Montreal, were statistically classified into seven optimal categories of exposure based on similarity in terms of active living potential measured by population density, land use mix, and geographical accessibility to proximity services. Classifications were then imported into a geographical information system resulting in the creation of 898 zones which were homogenous in terms of categories of exposure. The value of this approach was established by quantifying and comparing the degree of homogeneity of zones and census tracts in terms of active living potential. Then, area effects of active living potential on walking behaviours of 2716 adults (aged 45 years and older) were examined by comparing results obtained when areas were operationalised by the newly-defined zones rather than by census tracts.

Results: Although census tracts are homogenous with respect to accessibility to services, they are heterogeneous in terms of population density and land use mix. Furthermore, a majority of census tracts (55.5%) are characterised by a combination of at least three categories of exposure, thus by environments characterised by different levels of active living potential. Validity of census tracts for studying the environmental determinants of walking is therefore limited. Geographical accessibility to proximity services appears to be an important determinant of walking at both zone and census tract levels. Although effect size is larger at the zone-level, estimates are similar to those observed across census tracts. Nonetheless using area units homogenous in exposures increases the internal validity of the study design and therefore improves the soundness of statistical inference.

Conclusion: An approach that consists of designing spatial units maximising homogeneity of exposures is advantageous for defining areas as it strengthens internal validity without becoming unwieldy. Replication studies are required to further assess the value of the novel method for designing zones relevant in relation to other health indicators.

Keywords : Small-area analysis, residential characteristics, walking, multilevel models, geographic information systems.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

Article 1[a]

Article 3

[a] Les tableaux 2 et 3 sont présentés à l'Annexe I.

LISTE DES FIGURES

Introduction

Article 1

Article 2

Article 3

Discussion

LISTE DES ABRÉVIATIONS

Abréviations françaises

AD	Aires de diffusion
SR	Secteurs de recensement

Abréviations anglaises

CI	Confidence interval
DA	Dissemination areas
ERR	Event rate ratio
ICC	Intraclass correlation coefficient
GIS	Geographic information systems

À mes parents, à ma sœur

qui m'ont soutenue et encouragée tout au long de mon doctorat, et qui ont toléré, accepté et ri de mes absences et égarements.

Remerciements

La réalisation d'une thèse de doctorat ne peut s'accomplir sans le soutien et la collaboration de plusieurs personnes et organismes.

D'abord, je voudrais remercier et exprimer ma plus sincère reconnaissance à ma directrice Lise Gauvin et à mes co-directeurs de recherche, Jean-Marc Brodeur et Philippe Apparicio pour leur contribution exceptionnelle à ce projet de doctorat. Leur expertise, rigueur scientifique, générosité, curiosité et sens critique m'ont permis d'affiner ma pensée sur le rôle de l'espace pour la santé publique, de consolider ma passion pour la recherche, d'acquérir de nouveaux outils méthodologiques et de me surpasser pour mener à terme un projet de recherche dont je suis très fière.

Je tiens aussi à remercier Louise Potvin, Sarah Curtis et Kate Frohlich pour leurs judicieux conseils et pour avoir toujours eu le temps de discuter et d'explorer de nouvelles idées.

Un merci chaleureux à France Pinsonnault, Josée Tessier, Francine Auger, Jean-Guy Bouchard, Lorraine Trudeau et tout le personnel du Groupe de recherche interdisciplinaire en santé et du Département de médecine sociale et préventive pour leur support administratif et technique formidable.

À mes ami(e)s et collègues du programme de doctorat, ces dernières années n'auraient pas été aussi agréables sans les bonnes discussions lors des séminaires, dans les corridors et autour de la machine à café.

Finalement, je voudrais remercier les Instituts de recherche en santé du Canada, le programme de formation interdisciplinaire AnÉIS, le Groupe de recherche interdisciplinaire en santé, le Centre de recherche Léa-Roback sur les inégalités sociales de santé de Montréal et l'Équipe promotion de la santé de l'Université de Montréal pour leur soutien financier tout au long de mes études doctorales.

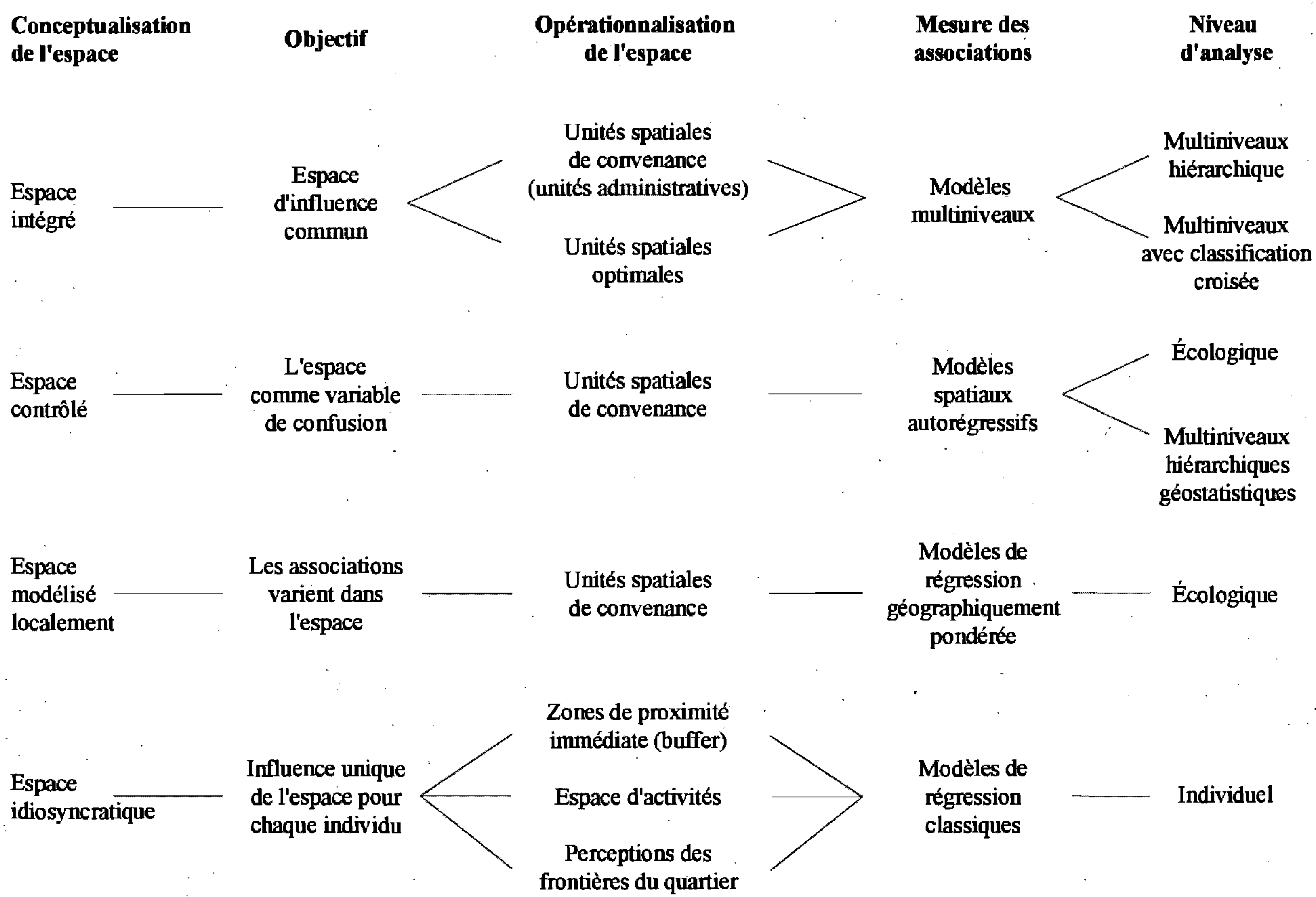
INTRODUCTION

L'importance des facteurs contextuels des milieux de vie pour la santé est reconnue depuis les origines de la tradition hippocratique de la médecine au 5^{e} siècle av. J.-C. À travers l'histoire de la santé publique, cette reconnaissance s'est maintenue et renforcée par les travaux d'éminents protagonistes tels John Graunt (1662), Louis René Villermé (1828), Edwin Chadwick (1842) et John Snow (1854) (Rosen, 1993). Ce n'est que depuis les années 1990, par contre, que la santé publique connaît une expansion considérable de travaux empiriques et théoriques portant sur le rôle des facteurs environnementaux pour la production et le maintien de la santé. Ces travaux s'inscrivent pour la plupart sous la nomenclature « effets de milieux »[1] sur la santé, et ont pour objectif l'étude des caractéristiques sociales et matérielles des milieux résidentiels en tant que facteurs de risque pour la santé (Berkman & Kawachi, 2000; Jones & Moon, 1993; Kearns, 1993; Kearns & Joseph, 1993; Macintyre, McIver, & Sooman, 1993). Cet objet de recherche résulte vraisemblablement de la convergence de deux tendances importantes. La première relève de l'appréciation des déterminants environnementaux de la santé et des multiples processus à travers lesquels ils influencent la santé, dont le contexte des milieux résidentiels. La seconde se rapporte à l'accessibilité de logiciels conviviaux qui permettent de réaliser des analyses statistiques avancées de types multiniveaux et spatiales.

Dans la formulation des problématiques de recherche en santé publique, les conceptualisations du milieu, ou de l'espace[2], sont multiples, chacune évoquant des objectifs, des définitions opérationnelles, des mesures et des niveaux d'analyse différents. Ces différentes conceptualisations sont illustrées à la Figure 1 et expliquées brièvement dans les paragraphes qui suivent.

[1] La terminologie « effets de milieux » est employée tout au long de cette thèse comme traduction de « *area effects* »; « milieu » réfère ainsi à « *area* ».
[2] Le terme « espace » est utilisé largement en référence à la conceptualisation du milieu.

Figure 1. Représentations de l'espace dans la recherche portant sur l'influence des milieux résidentiels sur la santé



L'espace *intégré* renvoie à la conceptualisation du milieu comme un niveau écologique d'influence dans lequel les individus sont imbriqués et exposés aux mêmes conditions. L'objectif est de déterminer si les facteurs contextuels des milieux, par exemple le statut socioéconomique, ont une influence statistiquement significative sur la santé des populations locales indépendamment des caractéristiques des individus (Ellen, Mijanovich, & Dillman, 2001; Pickett & Pearl, 2001; Riva, Gauvin, & Barnett, 2007). Par cette approche, le milieu est opérationnalisé par des unités spatiales administratives (par exemple les secteurs de recensement) ou optimales générées par diverses méthodes spatiales.

L'agencement spatial des données de santé et d'exposition (variables dépendantes et indépendantes) est tel que ces variables sont souvent autocorrélées c.-à-d. qu'elles partagent des valeurs similaires en fonction de leur proximité dans l'espace. Or, il est possible de tenir compte de cette autocorrélation spatiale dans les analyses statistiques. Par cette approche, l'espace est *contrôlé*; il est conceptualisé comme une variable de confusion (Chaix, Leyland, Sabel, Chauvin, Rastam, Kristersson et al., 2006; Chaix, Merlo, Subramanian, Lynch, & Chauvin, 2005). Les modèles spatiaux autorégressifs permettent d'intégrer l'autocorrélation spatiale des variables au niveau de la variable dépendante ou au niveau du terme d'erreur de l'équation de régression (Anselin, 1988; Anselin, Florax, & Rey, 2004; Banerjee, Carlin, & Gelfand, 2004). L'objectif de cette approche serait, par exemple, de vérifier si le statut socioéconomique du milieu de résidence, ajusté en fonction du statut socioéconomique des milieux avoisinants, influence ou non la santé.

Toutefois, l'autocorrélation spatiale des variables de santé et d'expositions donne lieu à un phénomène d'intérêt encore peu étudié en santé publique : la variation spatiale des effets de milieux (Anselin, 1988; Bailey & Gatrell, 1995; Fotheringham, Brundson, & Charlton, 1996). Il ne s'agit donc pas uniquement de contrôler l'autocorrélation spatiale, mais d'examiner l'instabilité spatiale des associations, notamment par l'emploi de modèles de régression géographiquement pondérée (*Geographically weighted regression*) (Fotheringham, Brunsdon, & Charlton, 2000, 2002). Par cette conceptualisation, l'espace est *modélisé localement*. Par opposition aux modèles de régression classiques, multiniveaux et autorégressifs qui produisent une équation prédictive pour l'ensemble de la région d'étude, la régression géographiquement pondérée produit une équation de régression pour chaque unité spatiale de la région à l'étude, par exemple pour chaque secteur de recensement (Cloutier, Apparicio, & Thouez, 2007; Nakaya, Fotheringham, Brunsdon, & Charlton, 2005). Ainsi, on pourrait investiguer si l'association entre le statut socioéconomique et la santé varie dans l'espace, car empiriquement, l'association peut s'avérer positive dans certains milieux, non significative dans d'autres, voire négative ailleurs.

Les conceptualisations précédentes de l'espace s'inscrivent dans une approche plus populationnelle dans la mesure où il est postulé que les individus d'un même milieu sont exposés aux mêmes facteurs environnementaux. Mais l'espace peut être *idiosyncratique*, c.-à-d. qu'il peut avoir une influence particulière pour chaque individu. Dans cette optique, le milieu peut être opérationnalisé de trois façons pour répondre à trois questions différentes. Le milieu peut être défini par l'environnement immédiat autour du lieu de résidence (Chaix et al., 2005; Frank, Schmid, Sallis, Chapman, & Saelens, 2005; Propper, Jones, Bolster, Burgess, Johnston, & Sarker, 2005; Wendel-Vos, Schuit, De Niet, Boshuizen, Saris, & Kromhout, 2004), délimité par exemple par une distance de 500 mètres; l'influence du statut socioéconomique sur la santé sera examinée dans cette zone de proximité. Une deuxième opérationnalisation correspond à l'espace vécu par l'individu, c.-à-d. aux divers lieux d'activité fréquentés quotidiennement (Takahashi, Wiebe, & Rodriguez, 2001; Wiles, 2003); l'influence sur la santé du statut socioéconomique de ces espaces d'activité pourrait être examinée. Finalement, l'espace perçu réfère à la perception que les gens ont des frontières de leur milieu de résidence et de ses caractéristiques (Coulton, Korbin, Chan, & Su, 2001); il s'agirait de déterminer si le statut socioéconomique du milieu de résidence, tel qu'il est perçu par les individus, influence la santé.

Les différentes conceptualisations de l'espace bien que complémentaires, répondent à des questions de recherches différentes. Pour la recherche et l'intervention en santé publique et en promotion de la santé, ces conceptualisations contribuent à la compréhension de la signification étiologique du milieu pour la santé et mettent l'accent sur les caractéristiques des milieux qui sont favorables ou délétères pour la santé (Macintyre & Ellaway, 2000). La majorité des études portant sur les effets de milieux, et notamment celles adoptant un devis d'analyse multiniveaux, épouse implicitement la conceptualisation de l'espace *intégré*. C'est sur cette conceptualisation de l'espace que la thèse s'est échafaudée et plus spécifiquement sur la problématique de la définition opérationnelle du milieu. Cette conceptualisation est privilégiée en raison de la reconnaissance des multiples niveaux écologiques pouvant influencer la santé et pour la considération conjointe des facteurs individuels et environnementaux dans l'explication des variations de santé.

L'espace intégré : la problématique de l'opérationnalisation du milieu

Les milieux résidentiels sont proximaux aux activités de la vie quotidienne. Pour cette raison, on présume qu'ils sont susceptibles d'influencer la santé des populations locales à travers les possibilités ou les obstructions qu'ils offrent pour l'adoption de saines habitudes de vie (Curtis & Jones, 1998; Macintyre, Ellaway, & Cummins, 2002). En continuité avec cette prémisse, de nombreuses études démontrent que, au-delà des caractéristiques individuelles, les facteurs contextuels des milieux ont une influence significative sur la santé des populations locales (Ellen et al., 2001; Pickett & Pearl, 2001; Riva et al., 2007).

Malgré ces résultats, l'opérationnalisation du milieu est problématique. En effet, alors que la fiabilité et la validité des mesures d'expositions écologiques (facteurs contextuels) s'améliorent (Cummins, Macintyre, Davidson, & Ellaway, 2005a; Frank et al., 2005; Oakes & Rossi, 2003; Pearce, Witten, & Bartie, 2006; Raudenbush & Sampson, 1999), il en est autrement en ce qui a trait à l'opérationnalisation du milieu.

En 1993, Jones et Moon observaient que, malgré l'intérêt certain porté au concept de milieu de vie :

> « *Seldom, however, does location itself play a real part in the analysis; it is the canvas on which events happen but the nature of the locality and its role in structuring health status and health-related behaviour is neglected* » (Jones and Moon, 1993, p.515).

Cette critique est toujours d'actualité; dans la majorité des études, le milieu est opérationnalisé par l'entremise d'unités spatiales administratives, tels les secteurs de recensement et les territoires de codes postaux, sans égard aux limites que ces opérationnalisations soulèvent pour l'étude des problématiques de santé (Coulton et al., 2001; Diez-Roux, 2000; Diez Roux, 2001; Gauvin, Robitaille, Riva, McLaren, Dassa, & Potvin, 2007; O'Campo, 2003; Pickett & Pearl, 2001; Subramanian, 2004; Subramanian, Jones, & Duncan, 2003). Il semble donc y avoir une certaine distanciation entre les construits que sont d'une part, le milieu et d'autre part, les expositions (ou le contexte). Par analogie, ces entités peuvent être perçues comme le contenant et le contenu. Ainsi, l'utilisation d'unités spatiales définies pour des motifs autres que pour l'étude des déterminants de la santé revient à traiter le milieu comme un contenant dénaturé d'un contenu bien apparié. Or, il existe un certain isomorphisme entre le milieu et son contexte de sorte qu'ils sont indissociables.

Plusieurs auteurs militent en faveur d'une opérationnalisation du milieu qui est fonction de la notion de « quartier », soit d'un espace localisé géographiquement et dont les frontières « naturelles » sont reconnues par les résidents (Coulton et al., 2001; Ross et al., 2004) et donc ayant une signification « écologique » (Pickett and Pearl, 2001). Mais qu'en est-il de la signification « étiologique » du milieu ? Les frontières des unités spatiales sont-elles cohérentes eu égard au phénomène de santé étudié (indicateur de santé et/ou exposition)? Les unités spatiales permettent-elles de mesurer sans erreur les facteurs d'exposition et leur influence sur la santé ? Soulevées par l'opérationnalisation du milieu, ces questions mettent en relief certains enjeux théoriques et empiriques.

Ainsi, cette thèse de doctorat porte sur la question suivante : *Quel est l'apport du développement et de l'application d'une approche novatrice pour créer des unités spatiales pour la compréhension des effets de milieux sur la santé ?* Cette approche est fondée sur la création de zones optimales homogènes au niveau d'une exposition spécifique associée à un indicateur de santé spécifique. Plus particulièrement, l'accent est mis sur le potentiel de vie active des milieux et les habitudes de marche.

Structure de la thèse

Afin de positionner la problématique de l'opérationnalisation du milieu, la thèse débute par une recension des écrits portant sur les effets de milieux sur la santé. Cette recension est présentée sous forme d'un article intitulé « *Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998* »; cet article a été publié dans le *Journal of Epidemiology and Community Health* (Riva et al., 2007). L'objectif de cet article est de déterminer l'étendue des résultats de recherche portant sur les déterminants sociaux et environnementaux de la santé et d'identifier les enjeux conceptuels et méthodologiques dans la mesure des effets de milieux sur la santé. Au total, 86 études employant des devis d'analyse multiniveaux ont été examinées; un tableau les synthétisant est présenté à l'Annexe I. La recension des écrits est suivie d'une discussion critique sur les enjeux liés à l'opérationnalisation du milieu. Une seconde (brève) recension des écrits fournit une synthèse de l'état des connaissances de l'influence des milieux résidentiels sur les comportements de marche.

Pour répondre plus directement à la question principale de la thèse, deux articles scientifiques ont été rédigés. Les données, mesures et analyses employées sont décrites en détail dans ces articles; pour cette raison, la thèse ne comprend pas de chapitre méthodologie proprement dit. Le premier article s'intitule « *Establishing the soundness of administrative spatial units for operationalising the active living* *potential of residential environments: an exemplar for designing optimal zones* » et a été accepté pour publication dans le *International Journal of Health Geographics*. Il vise à mettre au point une méthodologie pour créer des unités spatiales optimales, homogènes en termes de potentiel de vie active et d'établir la valeur de cette approche en quantifiant et comparant les degrés d'homogénéité des nouvelles unités et des secteurs de recensement au niveau du potentiel de vie active des milieux. Le second article s'intitule « *Studying the influence of area-level active living potential on walking: contribution of designing homogenous spatial units* » et examine l'influence des milieux sur la pratique de la marche en comparant les résultats obtenus lorsque le milieu est opérationnalisé par les nouvelles unités spatiales plutôt que par les secteurs de recensement. Cet article a été soumis à *Social Science and Medicine*. Suivant ces articles, le dernier chapitre de la thèse offre une discussion et une prise de position relativement à l'apport du développement et de l'utilisation d'unités spatiales homogènes au niveau des expositions pour la compréhension des effets de milieux sur la santé.

RECENSION DES ÉCRITS

ARTICLE 1

TOWARD THE NEXT GENERATION OF RESEARCH INTO SMALL AREA EFFECTS ON HEALTH: A SYNTHESIS OF MULTILEVEL INVESTIGATIONS PUBLISHED SINCE JULY 1998

Mylène Riva [1, 2, 3]
Lise Gauvin [1, 2, 3]
Tracie A Barnett [1, 2, 3]

[1] Département de médecine sociale et préventive, Université de Montréal, Canada

[2] GRIS, Groupe de recherche interdisciplinaire en santé, Université de Montréal, Canada

[3] Centre de recherche Léa-Roback sur les inégalités sociales de santé de Montréal, Canada

Article publié en octobre 2007 au volume 61 du Journal of Epidemiology and Community Health, aux pages 853 à 861.

Article 1. Rôle de chaque auteur

Riva, M., Gauvin, L., & Barnett, T. A. (2007). Toward the next generation of research into small area effects on health: A synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology and Community Health,* 61(10), 853-861.

Mylène Riva a conceptualisé la recension des écrits, codifié chacune des études et rédigé l'article.

Lise Gauvin a collaboré à la conceptualisation de l'étude et à la rédaction de l'article et à la validation d'une partie des résultats.

Tracie A. Barnett a collaboré à la validation d'une partie des résultats et à la rédaction de l'article.

ABSTRACT

In order to map out area effects on health research, this scoping study aims: (1) to inventory multilevel investigations of area effects on self-rated health, cardiovascular diseases and risk factors, and mortality among adults; (2) to describe and critically discuss methodological approaches employed and results observed; and (3) to formulate selected recommendations for advancing the study of area effects on health. Overall, 86 studies were inventoried. Although several innovative methodological approaches and analytical designs appeared in the literature, small areas are most often operationalised using administrative and statistical spatial units. Most studies used indicators of area SES derived from censuses and few provided information on the validity and reliability of measures of exposures. One consistent finding is that a significant portion of the variation in health is associated with area context independently of individual characteristics. Area effects on health, although significant in most studies, often depend on the health outcome studied, the measure of area exposure used, and the spatial scale at which associations are examined.

A brief search of extant literature on area effects on health shows a striking increase over the past decade in the number of studies adopting a multilevel approach to study social determinants of health. The impetus for such research likely results from a convergence of conceptual and methodological innovations including an appreciation of the importance of the social environment to health and greater accessibility of multilevel modelling techniques and software. However, multilevel investigations of area effects on health abound with conceptual and methodological challenges which have given rise to numerous debates. Debated issues are summarised in Table 1.

Table 1. Conceptual and methodological issues raised in the literature on area effects on health

Conceptual and methodological challenges	**Description**
Conceptualising causal pathways [2-10]	Absence of fully articulated theoretical frameworks and formulation of testable hypotheses.
Designating the ecologic unit of analysis [1, 11-12]	Ecologic units of analysis have been referred to as neighbourhood, small area, local area, and place. These labels have been used interchangeably without concern for differences in conceptual and operational definitions even though some have argued that there are substantive reasons for appropriately defining the nature of the ecological unit of analysis.
Defining the spatial contours of the ecological unit of analysis [1, 5, 8, 13-15]	Spatial contours of small areas are mostly delimited by existing administrative and statistical spatial units. However, these areas may be of limited utility in examining the association between area-level exposures and health outcomes because they may lack any intrinsic meaning in relation to health, they may not correspond to the spatial distribution of environmental features (ecologic exposures) associated with health, and they may be inconsistent with how residents define and experience their residential area.
Defining ecologic exposures [5, 7, 13, 15, 17-22]	Little attention has been devoted to conceptually and operationally defining ecologic exposures as researchers have tended to aggregate data from individuals to create meaningful area variables.
Controlling for individual-level variables [2, 5, 13, 15, 18, 23]	Lack of consistency in controlling for individual-level variables and further lack of consensus on whether individual variables should be conceptualised as confounders, moderators, or mediators of the associations between ecologic exposures and health outcomes.

Power, sample size, and representativeness [5, 13, 15, 25]	Usual considerations surrounding statistical power and sample size have been neglected.
Use of multilevel modelling techniques [4, 13, 15, 23-24, 26-28]	Appropriate use of multilevel modelling techniques have been applied sub-optimally thus limiting novel perspectives that might ensue from their judicious application: results are mainly reported for fixed effects whereas the potentials of discussing conceptual and methodological ramifications of random effects have been ignored.
Disentangling context from composition [7-8, 13, 18, 27, 29-34]	Some authors argue for disentangling the portion of the between area variation in health that is attributable to areas in which people live (contextual effect) from the portion attributable to individuals' characteristics (compositional effect) whereas others argued that this is a "false" issue as context and composition are inextricably intertwined.
Dearth of longitudinal and experimental studies [8, 18, 34-35]	Few studies have relied on research designs other than cross-sectional. This limits ascertaining the duration and timing of ecologic exposures, addressing selection bias, and ascribing causality. Furthermore, as people, areas, and the relationship between the two may change over time, using longitudinal designs is of accrued importance.

In a previous review of social determinants studies examining effects of area socioeconomic status (SES) on health, 23 studies out of 25 reported significant associations between at least one measure of area SES and health, while controlling for individual SES.[1] The authors concluded that data supported the existence of modest small area effects on health but that extant data were replete with methodological problems. More specifically, they stated that "it is clear from our review that investigations of the role of neighbourhood-level [small area] social factors on health are characteristics of preliminary, exploratory studies in epidemiology. Certain aspects of study design are in need of improvement before the field can advance [...] We hope that this review will show what has already been achieved and point the way to more sophisticated studies of societal determinants of health" (pp.120-121).

In an effort to map out multilevel research on social determinants of health, to identify the types of evidence available, and to gauge whether or not "more sophisticated studies" are being conducted, we undertook a scoping study of research of area effects on health published between July 1998 and December 2005. Unlike the more familiar systematic review, a scoping study addresses broad research topics where many different study designs are applied with the aim of comprehensively examining the extent, range, and nature of research activity and to identify key concepts and results.[36-37]

Given the broad diversity of studies, we restricted the scoping review to multilevel investigations of area effects on self-rated health (SRH), cardiovascular disease (CVD) and risk factors, and mortality among adults. These health indicators were selected because of their relevance to understanding the broader socio-spatial patterning of health. SRH is a highly predictive measure of morbidity and mortality, independent of other medical, behavioural, and/or psychosocial factors,[38] and CVD is one of the leading causes of mortality in developed countries.

We further restricted study selection to multilevel investigations allowing for estimation of between area variation (random effects). As pointed out by Merlo and colleagues,[24] "clustering of individual health within neighbourhoods [areas] is not a statistical nuisance that only needs to be considered for obtaining correct statistical estimations, but a key concept in social epidemiology that yields important information by itself (p. 443). Measures of variation, as they inform on the portion of health differences among people that may be attributable to the areas in which they live, are central to understand the significance of specific contexts for health.[24]

In keeping with the framework for conducting a scoping study proposed by Arksey and O'Malley,[36] the specific objectives of the scoping study were to: (1) inventory multilevel investigations of area effects on SRH, CVD and risk factors, and mortality; (2) describe and critically discuss methodological approaches employed and results observed; and (3) formulate selected recommendations for advancing the study of area effects on health.

METHODS FOR THE SCOPING STUDY

The scoping study involved several steps. First, identification of studies published between July 1998 and December 2005 ensued from a comprehensive search strategy using the Medline database. We first used "neighbourhood/neighbourhood or area" and "multilevel/multi-level or hierarchical" as words in the title or abstract of articles; this search yielded 634 entries. Inclusion criteria for studies were: publication in English-language peer-reviewed journals, data from adult populations in industrialised countries, use of a multilevel design with at least two units of analysis including individuals and areas, and measurement and analysis of health indicators at the individual level. Studies exploring between-country variation in health were excluded, though areas operationalised by sub-regions of a country, e.g. states, were included. Of the 634 studies, 67 meet the inclusion criteria.

We conducted additional searches using "contextual effects, place effects, census tracts, and community" as words in title or abstract, and "residence characteristics and small area analysis" as Medical Subject Headings (MeSH), again limiting the search to studies referring to multilevel methods in their title or abstract. Three additional studies were identified. References lists of all studies compiled and of other sources were reviewed, yielding an additional 16 studies. In all, eighty-six studies meeting inclusion criteria were retained. In two studies, authors reported findings for both self-rated health and CVD risk factors; these studies appear in both categories.[39-40] Overall the sample of the scoping study comprised 88 multilevel investigations of area effects on health.

In the second step, studies were coded by one author (MR) along the following dimensions: 1) citation and study location, 2) health indicator/analytic variable, 3) research design, year of data collection for individual sample, 4) individual sample size and sex/age distribution, 5) individual characteristics adjusted for, 6) area sample size and operational definition, 7) area-level exposures, 8) crude between area variation, 9) adjusted between area variation (for individual-level variables), and 10) summary of significant findings of adjusted area effects. The two other authors (LG and TAB) cross-validated half of the studies. Coding scheme and abbreviations are summarised in Table 2[3].

In the third step, in order to gauge the accuracy of data compiled, we established inter-author agreement in a random sample of about 25% of studies (n=21) where the coding of one of the authors (MR) was compared to that of another author (TAB) for all coding dimensions, except "citation/location" and "summary findings of area effects". For every dimension, each source of information was equated with one observation. Discrepancies in values reported were considered a disagreement. Overall inter-author agreement was 92.0% (43 disagreements out of 513 observations), with agreement ranging between 81.0% and 100% across dimensions. Finally, summary statistics were compiled using the total sample of investigations (n=88) as the unit of analysis.

RESULTS

Results of study coding appear in Table 3[3]. Studies are listed alphabetically by surname of first author within each category of health indicator, i.e., self-rated health (n=39), cardiovascular morbidity and risk factors (n=32), and mortality (n=17). Table 4 presents summary statistics for research design, operational definition of area contours, and exposure as a function of health indicator and time period.

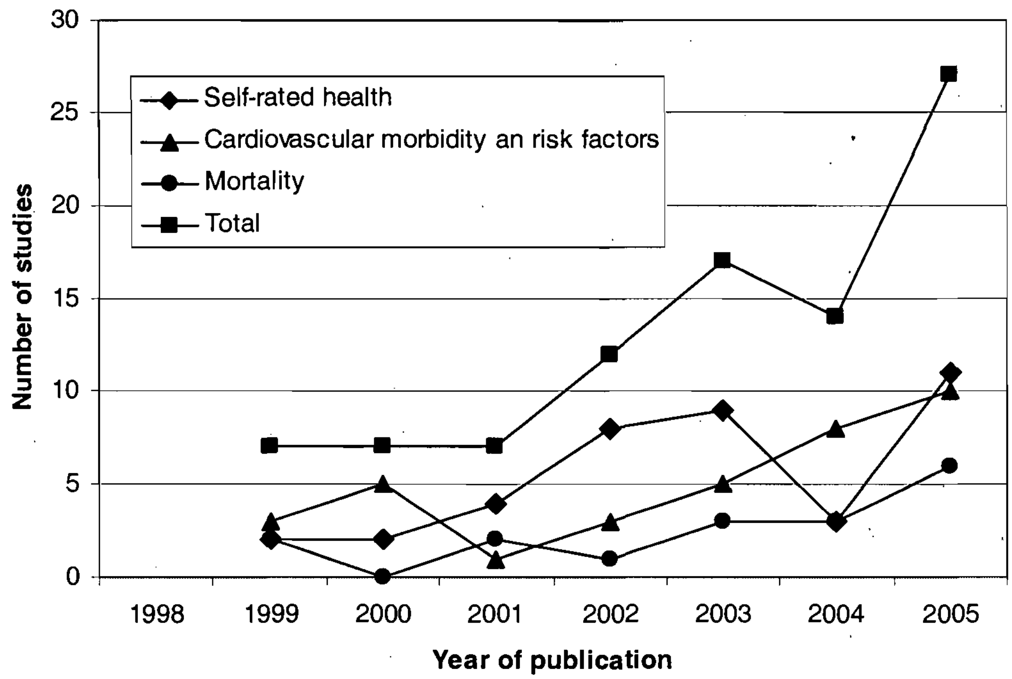
Year of publication and location of studies

There has been a marked increase in the number of studies published on area effects on self-rated health, CVD and risk factors, and mortality over the 1998-2005 time period, which almost doubled from 2004 to 2005 (Figure 1). Most results are from area effects examined in the United States (n=37) and the United Kingdom (n=14),

[3] Les tableaux 2 et 3 se trouvent à l'Annexe I

although several studies involved data collected in Canada (n=10), the Netherlands (n=8), and Sweden (n=8).

Figure 1. Trends in publication of multilevel investigations of area effects on self-rated health, cardiovascular morbidity and risk factors, and mortality



Research design and analytical variables

As shown in Table 4, a majority of studies (80.7%) had cross-sectional designs, whereas others adopted longitudinal designs (17.0%) wherein a majority of studies involved data from a cohort that were matched with vital statistics record to examine associations with mortality and CVD at a later time (designated as follow-up in Table 3). Linear multilevel models for continuous and logistic multilevel models for dichotomous outcomes were the most commonly used statistical models, although some analyses were performed on models for ordinal outcomes.[45, 48, 57, 64, 67, 76]

Table 4. Summary statistics for research design and operational definition of area contours and exposure as a function of health indicator and time period

	Self-rated health			Cardiovascular morbidity and risk factors			Mortality			TOTAL
Coding	**1998-2002**	**2003-2005**	**1998-2005**	**1998-2002**	**2003-2005**	**1998-2005**	**1998-2002**	**2003-2005**	**1998-2005**	**1998-2005**
dimensions	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Design										
Cross-sectional	16 (41.0)	23 (59.0)	39 (100.0)	10 (31.3)	17 (53.1)	27 (84.4)	2 (11.8)	3 (17.6)	5 (29.4)	71 (80.7)
Longitudinal	0 (0.0)	0 (0.0)	0 (0.0)	1 (3.1)	3 (9.4)	4 (12.5)	3 (17.6)	8 (47.1)	11 (64.7)	15 (17.0)
Case-Control	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (3.1)	1 (3.1)	0 (0.0)	1 (5.9)	1 (5.9)	2 (2.3)
Total	16 (41.0)	23 (59.0)	39 (100.0)	11 (34.4)	21 (65.6)	32 (100.0)	5 (29.4)	12 (70.6)	17 (100.0)	88 (100.0)
Area definition										
Administrative/ statistical	14 (35.9)	19 (48.7)	33 (84.6)	11 (34.4)	20 (62.5)	31 (96.9)	5 (29.4)	10 (58.8)	15 (88.2)	79 (89.8)
Other	2 (5.1)	4 (10.3)	6 (15.4)	0 (0.0)	1 (3.1)	1 (3.1)	0 (0.0)	2 (11.8)	2 (11.8)	9 (10.2)
Total	16 (41.0)	23 (59.0)	39 (100.0)	11 (34.4)	21 (65.6)	32 (100.0)	5 (29.4)	12 (70.6)	17 (100.0)	88 (100.0)
Area exposures										
Derived	15 (26,3)	20 (35,1)	35 (89.7)	8 (21.1)	17 (44.7)	25 (78.1)	5 (22.7)	12 (54.5)	17 (100.0)	77 (87.5)
Integral	5 (8,8)	13 (22,8)	18 (46.2)	3 (7.9)	8 (29.6)	11 (34.4)	2 (9.1)	3 (13.6)	5 (29.4)	34 (38.6)
None	1 (1,8)	3 (5,3)	4 (10.3)	0 (0.0)	2 (5.3)	2 (6.3)	0 (0.0)	0 (0.0)	0 (0.0)	6 (6.8)
Total*	21 (36,8)	36 (63,2)	39 (100.0)	11 (28.9)	27 (70.1)	32 (100.0)	7 (31.8)	15 (68.1)	17 (100.0)	88 (100.0)

*Percentages do not add up to 100% because some studies used both derived and integral measures of area exposures

Individual data: Sample size and variables

Sample size of individuals ranged between 577 [92] and 2 637 628,[101] with a median of 8 606 individuals. Sixteen percent of studies had a sample size over 100 000 individuals, but the majority of studies (61%) had a sample size under 10 000.

Most studies controlled for age, gender, SES, and marital status, but some controlled for other individual characteristics such as health-related behaviours, medical conditions, perception of area characteristics, social network, and years of residency in area. Seven studies did not control for individual socioeconomic status.[56, 103-104, 112, 116, 121, 124] Most studies targeted general populations, but some restricted their focus to men,[96, 188] older adults,[64, 88, 92] and racial/ethnic groups.[54-55, 59, 80, 83, 95]

Area data: Operational definition, sample size, and exposures

As shown in Table 4, the majority of studies (89.8%) operationalised areas using statistical (e.g. census tracts) and/or administrative spatial units (e.g. city-defined neighbourhoods, boroughs, local authorities). One study delimited areas using geographical information systems,[107] and others clustered statistical/administrative spatial units based on similarities in terms of socioeconomic status, demographic composition, and type of area.[41, 45-48, 76, 110, 116] Most studies had a two-level structure, with individuals nested within areas, though some had more complex structures, including cross-classification,[110] and three-level structures e.g., individuals nested within households within area(s) [63] or individuals nested within several hierarchically structured area units.[39, 44, 58, 85, 111, 119, 121]

Six studies did not report area-level sample size. Among studies for which data were reported, sample size ranged from 9 [56] to 12 344 areas.[111] Average within-area sample size ranged from 1 [100] to 36 387 individuals.[117] Half of the studies

(52.4%) had an average within area sample size of 50 individuals; for 10% of the studies, the within area sample size was less than 5 individuals.

Area-level indicators of socioeconomic status such as deprivation, education, and unemployment, were generally aggregates of individual-level variables derived from censuses and survey data. Others derived measures of area social context (e.g. social cohesion, social capital) by aggregating individuals' perceptions or by application of ecometric procedures.[45-46, 48] Over a third of studies (38.6%) operationalised area exposures using integral measures, i.e. features of areas only measurable at an ecological level (Table 4). The most commonly used integral measure was income inequality, but other studies relied on characteristics of the social and built environment,[e.g. 50, 67, 75, 92, 112] urban sprawl,[87] and availability of services and parks.[e.g. 81, 88, 92, 95, 106-107]

Summary of findings about area effects

Among 47 studies that reported on between area variation after adjusting for individual characteristics (minimally age, sex, and SES), twenty-seven (57.4%) studies reported significant between area variation for at least one subgroup (defined either by individual or area-level characteristics). Although several studies report one or the other, several studies did not report variance components for both unadjusted and adjusted models.

Of the 88 studies, six focussed on between area variation in health only [56, 60, 63, 73, 89, 94] and 82 studies examined main area effects. Of these studies, six did not report significant direct area effects on health.[51, 75, 93, 96, 120, 122] All other studies reported significant associations between at least one measure of area exposure and at least one health indicator. Significant cross-level interactions were observed, indicating that subgroups of individuals may be differentially influenced by certain area characteristics, and by interacting area characteristics. More specifically, effects of area deprivation on poor health, unhealthy behaviours, and risk of mortality were often greater among low SES individuals and among women.

Self-Rated Health

Thirty-nine studies examined area effects on self-rated health. Four studies focussed on between area variation only, and 35 were cross-sectional investigations of associations between area SES and SRH. In all but two studies,[51, 75] significant associations were observed between at least one measure of area SES and SRH. More specifically, less favourable area socioeconomic conditions were associated with poorer SRH. Area affluence, positive perceptions of area environment, and higher area levels of collective efficacy and social capital were predictive of better SRH, although area residential stability was associated with poorer health.[46, 48] One study reported associations between poor SRH and unfavourable area-level opportunity structures and social functioning, such as poor physical quality of residential environment, lower political engagement, and lower transport wealth.[50] Several studies reported significant effects of social processes, characteristics of the built environment, and perceptions of area characteristics as potential mediating pathways of the association between area deprivation/inequality and SRH.[46, 48, 51, 53, 76]

Cardiovascular morbidity and risk factors

Thirty-two studies were investigations of CVD and risk factors, of which two examined between area variation only.[89, 94] Some studies employed follow-up [96, 101-102] and case-control research designs,[99] but most were cross-sectional.

Twenty-three studies examined direct main effect of area deprivation on CVD and risk factors. In all studies, at least one measure of area deprivation was associated with greater risk for CVD and with at least one, but not all, of the risk factors investigated. In general, greater levels of area deprivation and inequality were associated with greater likelihood of unhealthy dietary habits, smoking, overweight and obesity, and physical inactivity. However, positive associations between affluence and smoking and drinking behaviours were also observed,[78-79, 104] and in one study, greater state inequality was associated with lower body mass index among white women.[80] Greater likelihood of walking was observed in more deprived areas,[88, 106] but also in more socially cohesive areas.[88]

In studies where area exposures were operationalised with integral variables, results showed that in less sprawling areas, risk of being overweight or obese was lower and levels of walking were higher.[87] Involvement in physical activity and walking was more likely in areas characterised by greater availability, accessibility, and density of selected services and green spaces.[87-88, 92, 106, 107] Dietary habits were associated with the presence of supermarkets and full-service restaurants in the area,[95] and greater mean distance to alcohol outlets was associated with lower alcohol consumption.[98] High convenience store density and lower distance to convenience store were associated with smoking, although this association was not significant in models controlling for area SES.[81]

Mortality

Seventeen studies examined area effects on mortality, most of which reported results from matching of cohort data with vital statistics.

In all but two studies,[120, 122] results showed that at least one measure of area SES was associated with all-cause and cause-specific mortality, such that greater area deprivation and income inequality were significantly associated with greater risk of mortality. When adjusting for area SES, greater area religious affiliation was associated with lower risk of all-cause mortality,[113] and areas with higher social capital were associated with lower risk of all-cause mortality and mortality from heart disease,[116] while lower levels of social cohesion were associated with higher risk of all-cause mortality.[118] Area SES confounded the association between air pollution and all-cause and cause-specific mortality.[115]

DISCUSSION

Results of the scoping study showed that the typical methodological approach for multilevel investigations of area effects remains a cross-sectional two-level study wherein individuals are nested within areas delimited by administrative/statistical spatial units, area-level indicators are operationalised using aggregates of individual variables, and direct associations between area exposures and individual-level outcomes are adjusted for selected individual characteristics. A number of large datasets produced several publications, and it should be noted that findings emanating from the same dataset clearly are not independent.

Over reliance on cross-sectional research designs raises the critical issue of "self-selection", i.e. the fact that people will be selected into residential areas based on individual attributes which are themselves related to health.[2] To deal with this issue, one frequently-used strategy consists of controlling for individual-level variables that are potential confounders of associations between area characteristics and health outcomes. Although useful, this modelling strategy does not overcome problems associated with misspecification resulting from omitted or mismeasured individual-level variables, or with lack of statistical power. Addressing these methodological issues rests on crafting longitudinal studies, assessing individual characteristics over the life course, and endeavouring to study cross-level interactions and mediating pathways. In this regard, several authors have underscored the value of innovative methodological approaches.[125]

In multilevel studies, there are at least two units of analysis, individuals and areas. Yet, the attention that is usually devoted to measuring individual attributes is infrequently carried over to areas. For example, in six studies, area sample size was not reported, and most studies provided little information on the validity and reliability of area-level measures of exposures. When measuring exposure, most studies relied on indicators of area SES derived from censuses and other surveys. Although easily accessible, such measures provide only truncated information about the context of areas,[7, 17] and may in fact be endogenous to the composition of the areas as they are determined by individual characteristics of residents.[2, 35] To this end, some studies have tapped into measuring area exposures by means of ecometric procedures.[20, 23, 45-46, 48]

Most studies employed administrative and/or statistical spatial units to define area contours despite recognised limitations, i.e. their potential lack of intrinsic meaning in relation to health.[1, 5, 8, 13-15] Such limitations are evidenced by variations in the strength and magnitude of area effects on health according to the operational definition of areas.[44, 58, 65, 111] We view as particularly innovative recent initiatives to define area contours by delimiting a radius around individual residential location and postcodes.[107, 126-127]

Between area variation was reported using a variety of statistical parameters including variance components and standard error coefficients, intraclass correlation coefficients for continuous and dichotomous outcomes, plausible value ranges, and others, which precludes comparisons across studies. Furthermore, several studies did not report variance components for either unadjusted or adjusted models. This is unfortunate given the importance of measures of variation for understanding the socio-spatial patterning of health. There is clearly a need for more detailed and consistent reporting of between area variation.[24]

Power estimation in multilevel studies is complicated by the need to account for the nested structure of the data.[128] Only one study reported power calculations.[91] Routine reporting of a priori power estimates and post hoc effect-size calculations is warranted in multilevel studies.

Although significant area effects were reported in a majority of studies critical examination of findings revealed inconsistencies within and across studies. For example, one study examined associations between area deprivation and SRH for different operational definitions of area; significant area effects were observed for one type of area only.[65] Others have examined effects of area deprivation on several CVD risk factors, but observed significant effects only for some of them.[78, 83] Within individual studies, associations between health and several indicators of area SES were frequently investigated. Although there is probably multicolinearity between measures of area exposures,[e.g. 44, 81, 119, 124] often only certain indicators were significantly associated with the health outcome. This suggests that indicators of area SES such as average income, educational attainment, and income inequality may be tapping into different aspects of the social environment and may be differently associated with specific health outcomes. In addition, a few studies reported no significant area effects. Of concern, findings of the scoping study show that area effects on a specific health indicator may be dependent of the measure of area exposure and the spatial level (area unit) at which associations are investigated. This clearly underscores the importance of conceptualising plausible causal pathways in the search for new knowledge.

The diversity in research designs employed and reported results calls attention to the need to move toward a set of reporting guidelines for multilevel investigations of area effects on health. This could include, but not be exclusive to 1) greater details accorded to the description of the nested structure of the database with systematic reporting of sample size at both individual and area levels, and a more explicit discussion of the validity and reliability of area-level measures of exposures; 2) description of the analytical strategy in such a way that would allow replication of analyses; and 3) better description of the between cluster variation in health outcomes, both in unadjusted and adjusted models for individual characteristics.[24]

Limitations

The main limitation of our scoping study pertains to inclusion criteria for studies. A first issue concerns keyword and Medical Subject Heading (MeSH) searches in the Medline database. Because the area of research emerged only recently, keywords listed by authors do not necessarily correspond to keywords used for searching the literature in the MeSH system. Rather, when searching for "neighbourhood" and "area" MeSH uses the terms "residence characteristics" and "small area analysis". Thus if these MeSH terms were not used to classify an article, and if the selected keyword, i.e. "neighbourhood", "area", "multilevel", and "hierarchical" were not specified either in the title or the abstract of an article, then it would not have been identified. To overcome this problem, more encompassing terms were used and references lists of all identified articles were reviewed to identify studies. Nonetheless, we acknowledge that some studies may not have been identified with our search strategy.

A second limitation pertains to the exclusion of the literature on area effects on mental health, non-CVD morbidity, paediatric populations, maternal health, and health services utilisation. Several of these studies have used innovative methodological approaches and analytical procedures, which could provide the reader with other perspectives on area effects on health.[126-127, 129-130] It seems relevant to undertake a review of these studies as well to ascertain similarities and differences in the range and types of investigations conducted.

CONCLUSION

Results of the scoping study critically raise several issues. One issue pertains to whether or not "true" area effects are concealed by less than adequate methodologies. This may indeed be the case. Several studies showed significant between area variation and area effects independently of individual characteristics. These are quite considerable findings given that most studies suffered from methodological limitations and lack of precision when operationalising and measuring context. Significant results garnered with such measurement error probably underestimate effect sizes.

In recent years, increased attention has been directed towards formulating and testing theoretically based pathways between more specific area-level measures of exposures and more specific health indicators, such as the effect of the density of food stores on dietary habits, or the effect of urban form on physical activity involvement. These measures of area exposures may be more proximal to influence everyday health-related behaviours, and thus operating on the pathways between area SES and broader health outcomes.

A second issue is whether or not widely used methodologies for studying areas and health are well-suited to the task. Multilevel modelling, without being a panacea, is a suitable statistical procedure that can be used to analyse data with nested sources of variability, while accounting for the non independence of within cluster observations, i.e. addressing the non-random processes situating people with similar characteristics into certain type of areas.[25] As operationalisation of area contours essentially falls back on using readily available spatial units such as census tracts, space is fragmented into seemingly independent area units, therefore ignoring spatial associations between areas. As pointed out by Chaix and colleagues [126] multilevel modelling procedures are based on the assumption that spatial correlation can be reduced to within-area correlation. For this reason, multilevel approaches may provide only limited information on the spatial distribution of health outcomes and ecological exposures, both when modelling variations and investigating associations.[126] Further debates underscore limitations of multilevel models in detecting causal effects of area exposures on health outcomes.[2, 15, 35, 131].

Another issue relates to the most promising approaches in this area of research. In recent years, didactical and conceptual tutorials linking social epidemiological concepts to multilevel analysis have been published.[24, 26, 28, 132] Innovative methodological approaches and analytical procedures have been applied to examine area effects on health, such as geographical information systems used to operationalise area contours and ecological exposures,[107, 126, 127] development of measurement technology, e.g. ecometrics,[20, 22, 45-46, 48] geostatistical modelling and spatial analysis,[47, 115, 126, 133,] multilevel path analyses,[92] and multilevel structural equation modelling to test potential mediating pathways between area exposures and health outcomes.[53, 88] Others have underscored the value of developing experimental research designs such as randomised community trials,[35] and natural [134] and social experiments.[135]

What seems to emerge from the accumulating evidence on area effects on health is a "specific" research agenda. As argued by other authors,[17, 133, 136] we espouse the view that the adoption of a specific research approach to examine area effects on health, i.e. one that would conceptualise, operationalise, and measure associations between specific health outcomes and specific area exposures, across specific spatial area units may yield more informative evidence of area effects. Adopting a specific approach shows the greatest promise for advancing theoretically-based pathways, providing a basis for more precise definitions and measures of ecological exposures, and improved delimitations of area contours.

Acknowledgment

MR is the recipient of a Canada Graduate Scholarships Doctoral Award and TAB is recipient of a Postdoctoral Fellowship, both from the Canadian Institutes of Health Research. Both are working under the supervision of LG.

The authors would like to thank Steven Cummins for his helpful comments on an earlier draft of the manuscript.

What is already known

A significant portion of the variation in health is associated with area context independently of individual characteristics.

What this paper adds

This paper is a scoping study of 86 multilevel investigations of area effects on self-rated health, cardiovascular morbidity and risk factors, and mortality among adults.

Area effects on health, although significant in most studies, often depend on the health outcome studied, the measure of area exposure used, and the spatial scale at which associations are examined. This highlights the importance conceptualising causal pathways linking area context to health outcomes and devising appropriate methodological strategies in future research.

Policy implications

Various health outcomes are influenced by area context although the specific processes through which such influences occur remain unclear. The implementation and evaluation of policy interventions aimed at changing area exposures represents an opportunity to fill this knowledge gap.

References

1. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 2001;**55**:111-122.
2. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc Sci Med* 2004;**58**:1953-1960.
3. Kaplan GA. What's wrong with social epidemiology, and how can we make it better? *Epidemiol Rev* 2004;**26**:124-135.
4. Bingenheimer JB. Multilevel models and scientific progress in social epidemiology. *J Epidemiol Communiy Health* 2005;**59**:438-439.
5. Diez Roux AV. Multilevel analysis in public health research. *Annual Review of Public Health* 2000;**21**:171-192.
6. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int J Epidemiol* 2001;**30**:668-677.
7. Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise and measure them? *Soc Sci Med* 2002;**55**:125-139.
8. O'Campo P. Invited commentary: advancing theory and methods for multilevel models of residential neighborhoods and health. *Am J Epidemiol* 2003;**157**:9-13.
9. Susser M, Susser E. Choosing a future for epidemiology: I. Eras and paradigms. *Am J Public Health* 1996;**86**:668-673.
10. Susser M, Susser E. Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *Am J Public Health* 1996;**86**:674-677.
11. Berkman LF, Kawachi, I. *Social epidemiology*. New York, NY: Oxford University Press 2000.
12. Galster G. On the nature of neighbourhood. *Urban Stud* 2001;**38**:2111-2124.
13. Subramanian SV, Jones K, Duncan C. Multilevel methods for public health research. In Kawachi I, Berkman LF, eds. *Neighborhoods and health*. New York: Oxford University Press 2003:65-111.

14. Coulton CJ, Korbin J, Chan T, *et al.* Mapping resident's perceptions of neighborhood boundaries: a methodological note. *American Journal of Community Psychology* 2001;**29**:371-383.

15. Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med* 2004;**58**:1961-1967.

16. Harris RJ, Longley PA. Creating small measures of urban deprivation. *Environ Plan A* 2002;**34**:1073-1093.

17. Cummins S, Macintyre S, Davidson S, *et al.* Measuring neighbourhood social and material context: generation and interpretation of ecological data from routine and non-routine sources. *Health Place* 2005;**11**:249-260.

18. Diez Roux AV Investigating neighborhood and area effects on health. *American Journal of Public Health* 2001;**91**:1783-1789.

19. Diez Roux AV. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiol Rev* 2004;**26**: 104-111.

20. Gauvin L, Richard L, Craig CL, *et al.* From walkability to active living potential. An "ecometric" validation study. *Am J Prev Med* 2005;**28**(Suppl.2):126-133.

21. Macintyre S, Ellaway A. Ecological approaches: rediscovering the role of the physical and social environment. In: Berkman LF, Kawachi I, eds. *Social epidemiology*. New York: Oxford University Press 2000:332-348.

22. Raudenbush SW, Sampson RJ. Ecometrics: toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods. *Sociol Methodol* 1999;**29**:1-41.

23. Sampson RJ, Morenoff JD, Gannon-Rowley T. Assessing "neighborhood effects": social processes and new directions in research. *Annual Review of Sociology* 2002;**28**:443-478.

24. Merlo J, Chaix B, Yang M, *et al.* A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health* 2005;**59**;443-449.

25. Snijders TAB, Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modelling*. 1st ed. Thousand Oaks, CA: Sage 1999.

26. Merlo J, Yang M, Chaix B, *et al.* A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J Epidemiol Community Health* 2005;**59**;729-736.

27. Bingenheimer JB, Raudenbush SW. Statistical and substantive inferences in public health: issues in the application of multilevel models. *Annu Rev Public Health* 2004;**25**:53-77.

28. Merlo J, Chaix B, Yang M, *et al.* A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *J Epidemiol Community Health* 2005;**59**;1022-1029.

29. Blakely TA, Woodward AJ. Ecological effects in multi-level studies. *J Epidemiol Community Health* 2000;**54**:367-374.

30. Curtis S, Jones IR. Is there a place for geography in the analysis of health inequalities? *Sociol Health Illn* 1998;**20**:645-672.

31. Diez-Roux AV. The examination of neighbourhood effects on health: conceptual and methodological issues related to the presence of multiple levels of organisations. In: Kawachi I, Berkman LF, eds. *Neighborhoods and health.* New York: Oxford University Press 2003:25-64.

32. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multi-level models in health research. *Soc Sci Med* 1998;**46**:97-117.

33. Macintyre S, Ellaway A. Neighborhoods and health: an overview. In: Kawachi I, Berkman LF, eds. *Neighborhoods and Health.* New York: Oxford University Press 2003:20-42.

34. Yen IH, Syme SL. The social environment and health: a discussion of the epidemiologic literature. *Annu Rev Public Health* 1999;**20**:287-308.

35. Oakes JM. The (mis)estimation of neighbourhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 2004;**58**:1929-1952.

36. Arksey H, O'Malley L. Scoping Studies: towards a methodological framework. *International Journal of Social Research Methodology* 2005;**8**:19-32.

37. Mays N, Roberts E, Popay J. Synthesising research evidence. In: Fulop N, Allen P, Clarke A, Black N, eds. *Studying the organisation and delivery of health services: research methods.* London: Routledge 2001.

38. Idler E, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav* 1997;**38**:21-37.

39. Reijneveld SA. Neighbourhood socioeconomic context and self reported health and smoking: a secondary analysis of data on seven cities. *J Epidemiol Community Health* 2002;**56**;935-42.

40. Stafford M, Marmot M. Neighbourhood deprivation and health: does it affect us all equally? *Int J Epidemiol* 2003;**32**:357-366.

41. Béland F, Birch S, Stoddart G. Unemployment and health: contextual-level influences on the production of health in populations. *Soc Sci Med* 2002;**55**:2033-2052.

42. Blakely TA, Kennedy BP, Glass R, *et al.* What is the lag time between income inequality and health status? *J Epidemiol Community Health* 2000;**54**:318-319.

43. Blakely TA, Kennedy BP, Kawachi I. Socioeconomic Inequality in Voting Participation and Self-Rated Health. *Am J Public Health* 2001;**91**:99-104.

44. Blakely TA, Lochner K, Kawachi I. Metropolitan area income inequality and self-rated health - a multi-level study. Soc Sci Med 2002:**54**:65-77.

45. Browning CR, Cagney KA. Neighborhood structural disadvantage, collective efficacy, and self-rated physical health in an urban setting. *J Health Soc Behav* 2002;**43**:383-99.

46. Browning CR, Cagney KA. Moving beyond poverty: neighborhood structure, social processes, and health. *J Health Soc Behav* 2003;**44**:552-71.

47. Browning CR, Cagney KA, Wen M. Explaining variation in health status across space and time: implications for racial and ethnic disparities in self-rated health. *Soc Sci Med* 2003;**57**:1221-35.

48. Cagney KA, Browning CR, Wen M. Racial disparities in self-rated health at older ages: what difference does the neighbourhood make? *J Gerontol* 2005;**60B**:S181-90.

49. Craig N. Exploring the generalisability of the association between income inequality and self-assessed health. *Soc Sci Med* 2005;**60**:2477-488.

50. Cummins S, Stafford M, Macintyre S, *et al.* Neighbourhood environment and its association with self rated health: evidence from Scotland and England. *J Epidemiol Community Health* 2005;**59**:207-13.

51. Drukker M, van Os J. Mediators of neighbourhood socioeconomic deprivation and quality of life. *Soc Psychiatry Psychiatr Epidemiol* 2003;**38**:698-706.

52. Drukker M, Kaplan C, van Os J. Residential instability in socioeconomically deprived neighbourhoods, good or bad? *Health Place* 2005;**11**: 121-129.

53. Franzini L, Caughy M, Spears W, *et al.* Neighborhood economic conditions, social processes, and self-rated health in low-income neighborhoods in Texas: a multilevel latent variables model. *Soc Sci Med* 2005;**61**:1135-1150.

54. Gee GC. A multilevel analysis of the relationship between institutional and individual racial discrimination and health status. *Am J Public Health* 2002;**92**:615-23.

55. Gee GC, Takeuchi DT. Traffic stress, vehicular burden and well-being: a multilevel analysis. *Soc Sci Med* 2004;**59**:405-14.

56. Hopman WM, Berger C, Joseph L, *et al.* Is there regional variation in the SF-36 scores of Canadian adults? *Can J Public Health* 2002;**93**:233-237.

57. Hou F, Chen J. Neighbourhood low income, income inequality and health in Toronto. *Health Reports* 2003;**14**:21-34.

58. Hou F, Myles J. Neighbourhood inequality, neighbourhood affluence and population health. *Soc Sci Med* 2005;**60**:1557-569.

59. Karlsen S, Nazroo JY, Stephenson R. Ethnicity, environment and health: putting ethnic inequalities in health in their place. *Soc Sci Med* 2002;**55**:1647-661.

60. Lindstrom M, Moghaddassi M, Merlo J. Individual self-reported health, social participation and neighbourhood: a multilevel analysis in Malmo, Sweden. *Prev Med* 2004;**39**:135-41.

61. Malmstrom M, Sundquist J, Johansson SE. Neighborhood environment and self-reported health status: a multilevel analysis. *Am J Public Health* 1999;**89**:1181-86.

62. McCulloch A. Ward-level deprivation and individual and economic outcomes in the British Household Panel Survey. *Environ Plan A* 2001;**33**:667-684.

63. Pampalon R, Duncan C, Subramanian SV, *et al.* Geographies of health perception in Québec: a multilevel perspective. *Soc Sci Med* 1999;**48**:1483-490.

64. Patel KV, Eschbach K, Rudkin LL, *et al.* Neighborhood context and self-rated health in older Mexican Americans. *Ann Epidemiol* 2003;**13**;620-28.

65. Reijneveld SA, Verheij RA, Bakker DH. The impact of area deprivation on differences in health: does the choice of the geographical classification matter? *J Epidemiol Community Health* 2000;**54**:306-13.

66. Robert SA, Lee KY. Explaining race differences in health among older adults: the contribution of community socioeconomic context. *Res Aging* 2002;**24**:654-683.

67. Stafford M, Bartley M, Mitchell R, *et al.* Characteristics of individuals and characteristics of areas: investigating their influence on health in the Whitehall II study. *Health Place* 2001;**7**:117-129.

68. Stafford M, Martikainen P, Lahelma E, *et al.* Neighbourhoods and self rated health: a comparison of public sector employees in London and Helsinki. *J Epidemiol Community Health* 2004;**58**:772-78.

69. Stafford M, Cummins S, Macintyre S, *et al.* Gender differences in the associations between health and neighbourhood environment. *Soc Sci Med* 2005;**60**:1681-692.

70. Subramanian SV, Kawachi I, Kennedy BP. Does the state you live in make a difference? Multilevel analysis of self-rated health in the US. *Soc Sci Med* 2001;**53**:9-19.

71. Subramanian SV, Kim DJ, Kawachi I. Social trust and self-rated health in US communities: a multilevel analysis. *J Urban Health* 2002;**79(Suppl. 4)**:S21-S34.

72. Subramanian SV, Kawachi I. The association between state income inequality and worse health is not confounded by race. *Int J Epidemiol* 2003;**32**:1022-1028.

73. Subramanian SV, Kim D, Kawachi I. Covariation in the socioeconomic determinants of self rated health and happiness: a multivariate multilevel analysis of individuals and communities in the USA. *J Epidemiol Community Health* 2005;**59**:664-69 .

74. Subramanian SV, Acevedo-Garcia D, Osypuk TL. Racial residential segregation and geographic heterogeneity in black/white disparity in poor self-rated health in the US: a multilevel statistical analysis. *Soc Sci Med* 2005;**60**:1667-1679.

75. Veenstra G. Location, location, location: contextual and compositional health effects of social capital in British Columbia, Canada. *Soc Sci Med* 2005;**60**:2059-2071.

76. Wen M, Browning CR, Cagney KA. Poverty, affluence, and income inequality: neighborhood economic structure and its implications for health. *Soc Sci Med* 2003;**57**:43-860.

77. Xi G, McDowell I, Nair R, *et al*. Income inequality and health in Ontario: a multilevel analysis. *Can J Public Health* 2005;**96**:206-11.

78. Chaix B, Chauvin P. Tobacco and alcohol consumption, sedentary lifestyle and overweightness in France: a multilevel analysis of individual and area-level determinants. *Eur J Epidemiol* 2003;**18**:531-38.

79. Chaix B, Guilbert P, Chauvin P. A multilevel analysis of tobacco use and tobacco consumption levels in France. Are there any combination risk groups? *Eur J Public Health* 2004;**14**:186–90.

80. Chang VW, Christakis NA. Income inequality and weight status in US metropolitan areas. *Soc Sci Med* 2005;**61**:83-96.

81. Chuang YC, Cubbin C, Ahn D, *et al*. Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. *J Epidemiol Community Health* 2005;**59**:568-73.

82. Cubbin C, Winkleby MA. Protective and harmful effects of neighborhood-level deprivation on individual-level health knowledge, behavior changes, and risk of coronary heart disease. *Am J Epidemiol* 2005;**162**:559-68.

83. Diez-Roux AV, Nieto FJ, Caulfield L, *et al.* Neighbourhood differences in diet: the Atherosclerosis Risk in Communities (ARIC) Study. *J Epidemiol Community Health* 1999;**53**:55-63.

84. Diez-Roux AV, Link BG, Northridge ME. A multilevel analysis of income inequality and cardiovascular disease risk factors. *Soc Sci Med* 2000;**50**:673-87.

85. Duncan C, Jones K, Moon G. Smoking and deprivation: are there neighbourhood effects? *Soc Sci Med* 1999;**48**:497-505.

86. Ecob R, Macintyre S. Small area variations in health related behaviours; do these depend on the behaviour itself, its measurement, or on personal characteristics? *Health Place* 2000;**6**:261-74.

87. Ewing R, Schmid TL, Killingsworth R, *et al.* Relationship between urban sprawl and physical activity, obesity, and morbidity. *Am J Health Promot* 2003;**18**:47-57.

88. Fisher KJ, Li F, Michael Y, *et al.* Neighborhood-level influences on physical activity among older adults: a multi-level analysis. *J Aging Phys Act* 2004;**12**:45-63.

89. Kairouz S, Nadeau L, Siou GL. Area variations in the prevalence of substance use and gambling behaviours and problems in Québec: a multilevel analysis. *Can J Psychiatry* 2005;**50**:591-98.

90. Kavanagh AM, Goller JL, King T, *et al.* Urban area disadvantage and physical activity: a multilevel study in Melbourne, Australia. *J Epidemiol Community Health* 2005;**59**;934-940.

91. Leyland AH. Socioeconomic gradients in the prevalence of cardiovascular disease in Scotland: the roles of composition and context. *J Epidemiol Community Health* 2005;**59**:799-803.

92. Li F, Fisher KJ, Brownson RC, *et al.* Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *J Epidemiol Community Health* 2005;**59**:558-64.

93. Lindstrom M, Moghaddassi M, Merlo J. Social capital and leisure time physical activity: a population based multilevel analysis in Malmo, Sweden. *J Epidemiol Community Health* 2003;**57**:23-28.

94. Lindstrom M, Moghaddassi M, Bolin K, *et al.* Social participation, social capital and daily tobacco smoking: a population-based multilevel analysis in Malmo, Sweden. *Scand J Public Health* 2003;**31**:444-50.

95. Morland K, Wing S, Diez Roux A. The contextual effect of the local food environment on residents' diets: the Atherosclerosis risk in community study. *Am J Public Health* 2002;**92**:1761-1767.

96. Morris RW, Whincup PH, Lampe FC, *et al.* Geographic variation in incidence of coronary heart disease in Britain: the contribution of established risk factors. *Heart* 2001;**86**;277-283.

97. Robert SA, Reither EN. A multilevel analysis of race, community disadvantage, and body mass index among adults in the US. *Soc Sci Med* 2004;**59**:2421-2434.

98. Scribner RA, Cohen DA, Fisher W. Evidence of a structural effect for alcohol outlet density: a multilevel analysis. *Alcohol Clin Exp Res* 2000;**24**:188-95.

99. Stjärne MK, de Leon AP, Hallqvist J. Contextual effects of social fragmentation and material deprivation on risk of myocardial infarction: results from the Stockholm Heart Epidemiology Program (SHEEP). *Int J Epidemiol* 2004;**33**:732-741.

100. Sundquist J, Malmstrom M, Johansson SE. Cardiovascular risk factors and the neighbourhood environment: a multilevel analysis. *Int J Epidemiol* 1999;**28**:841-45.

101. Sundquist K, Malmstrom M, Johansson SE. Neighbourhood deprivation and incidence of coronary heart disease: a multilevel study of 2.6 million women and men in Sweden. *J Epidemiol Community Health* 2004;**58**:71-77.

102. Sundquist K, Winkleby M, Ahlen H, *et al.* Neighborhood socioeconomic environment and incidence of coronary heart disease: a follow-up study of 25,319 women and men in Sweden. *Am J Epidemiol* 2004;**159**:655-62.

103. Tonne C, Schwartz J, Mittleman M, *et al.* Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods *Circulation* 2005;**111**:3063-070.

104. Twigg L, Moon G, Jones K. Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators. *Soc Sci Med* 2000;**50**:1109-120.

105. van Lenthe FJ, Mackenbach JP. Neighbourhood deprivation and overweight: the GLOBE study. *Int J Obes* 2002;**26**:234-240.

106. van Lenthe FJ, Brug J, Mackenbach J. Neighbourhood inequalities in physical inactivity: the role of neighbourhood attractiveness, proximity to local facilities and safety in the Netherlands. Soc Sci Med 2005;**60**:763-75.

107. Wendel-Vos GCW, Sschuit AJ, De Niet R, *et al.* Factors of the Physical Environment Associated with Walking and Bicycling. *Med Sci Sports Exerc* 2004;**36**:725-30.

108. Borrell C, Rodriguez M, Ferrando J, *et al.* Role of individual and contextual effects in injury mortality: new evidence from small area analysis. *Inj Prev* 2002;**8**:297-302.

109. Bosma H, van de Mheen HD, Borsboom GJJM, *et al.* Neighborhood Socioeconomic Status and All-Cause Mortality. *Am J Epidemiol* 2001;**153**:363-371.

110. Curtis S, Southall H, Congdon P, *et al.* Area effects on health variation over the life-course: analysis of the longitudinal study sample in England using new data on area of residence in childhood. *Soc Sci Med* 2004;**58**:57-74.

111. Franzini L, Spears W. Contributions of social context to inequalities in years of life lost to heart disease in Texas, USA. *Soc Sci Med* 2003;**57**:1847-61.

112. Hembree C, Galea S, Ahern J, *et al.* The urban built environment and overdose mortality in New York City neighbourhoods. *Health Place* 2005;**11**:147-56.

113. Jaffe DH, Eisenbach Z, Neumark YD, *et al.* Does living in a religiously affiliated neighborhood lower mortality? *Ann Epidemiol* 2005;**15**:804–10.

114. Jaffe DH, Eisenbach Z, Neumark YD, *et al.* Individual, household and neighborhood socioeconomic status and mortality: a study of absolute and relative deprivation. *Soc Sci Med* 2005;**60**:989-97.

115. Jerrett M, Burnett RT, Ma R, *et al.*, Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiol* 2005;**16**:727-736.

116. Lochner KA, Kawachi I, Brennan RT, *et al.* Social capital and neighborhood mortality rates in Chicago. *Soc Sci Med* 2003;**56**:1797-805.

117. Marinacci C, Spadea T, Biggeri A, *et al.* The role of individual and contextual socioeconomic circumstances on mortality: analysis of time variations in a city of North West Italy. *J Epidemiol Community Health* 2004;**58:**199-207.

118. Martikanen P, Kauppinen TM, Valkonen T. Effect of the characteristics of neighbourhoods and the characteristics of people on cause specific mortality: a register based follow up study of 252 000 men. *J Epidemiol Community Health* 2003;**57**:210-17.

119. Mohan J, Twigg L, Barnard S, *et al.* Social capital, geography and health: a small-area analysis for England. *Soc Sci Med* 2005;**60**:1267-283.

120. Roos LL, Magoon J, Gupta S, *et al.* Socioeconomic determinants of mortality in two Canadian provinces: multilevel modelling and neighborhood context. *Soc Sci Med* 2004;**59**:1435-47.

121. Subramanian SV, Chen JT, Rehkopf DH, *et al.* Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in massachusetts. *Am J Public Health* 2005;**95**:260-65.

122. Veugelers PJ, Yip AM, Kephart G. Proximate and contextual socioeconomic determinants of mortality: multilevel approaches in a setting with universal health care coverage. *Am J Epidemiol* 2001;**154**:725-732.

123. Waitzman NJ, Smith KR, Stroup A. The direct and indirect effects of metropolitan area inequality on mortality a hierarchical analysis. *Ann N Y Acad Sci* 1999;**896**:347-49.

124. Yen IH, Kaplan GA. Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study. *Am J Epidemiol* 1999;**149**:898-907.

125. Oakes JM, Johnson PH. Propensity score matching for social epidemiology. In Oakes JM, Kaufman JS, eds. *Methods in social epidemiology*. San Francisco: Jossey-Bass 2006:364-386.

126. Chaix B, Merlo J, Subramanian SV, *et al.* Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *Am J Epidemiol* 2005;**162**:171-82.

127. Propper C, Jones K, Bolster A, *et al.* Local neighbourhood and mental health: evidence from the UK *Soc Sci Med* 2005;**61**:2065-083.

128. Raudenbush SW, Bryk A. *Hierarchical linear models: applications and data analysis methods*. 2nd ed. Newbury Park, CA: Sage 2002.

129. Frohlich KL. Potvin L, Gauvin L, *et al.* Youth smoking initiation: disentangling context from composition. *Health Place* 2002;**8**:155-166.

130. Drukker M, Kaplan C, Feron F, *et al.* Children's health-related quality of life, neighbourhood socio-economic deprivation and social capital: a contextual analysis. *Soc Sci Med* 2003;**57**:825-841.

131. Oakes MJ. Causal inference and the relevance of social epidemiology. *Soc Sci Med* 2004;**58**:1969-1971

132. Merlo J, Chaix B, Ohlsson H, *et al.* A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health* 2006;**60**:290–297.

133. Chaix B, Rosvall M, Lynch J, *et al.* Disentangling contextual effects on cause-specific mortality in a longitudinal 23-year follow-up study: impact of population density or socioeconomic environment? *Int J Epidemiol.* Advance Access published on February 1, 2006; doi:10.1093/ije/dyl009

134. Cummins S, Pettirew M, Higgins C, *et al.* Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. *J Epidemiol Community Health* 2005;**59**:1035-1040.

135. Levental T, Brooks-Gunn J. Moving to Opportunity: an experimental study of neighborhood effects on mental health. *Am J Public Health* 2003;**93**:1576-1582

136. Galea S, Ahern J. Invited commentary: Considerations about specificity of associations, causal pathways, and heterogeneity in multilevel thinking. Am J Epidemiol 2006;**163**:1079-1082.

Opérationnaliser le milieu : défis théoriques et empiriques

Tel qu'il a été démontré par les résultats des études multiniveaux recensées, une portion significative de la variation de la santé peut être attribuée aux facteurs contextuels des milieux résidentiels (Riva et al., 2007).[4] Or, certaines études indiquent que la force et la taille des effets de milieux sur la santé varient en fonction de la définition opérationnelle des milieux (Blakely, Lochner, & Kawachi, 2002; Cockings & Martin, 2005; Franzini & Spears, 2003; Hou & Myles, 2005; Krieger, Chen, Waterman, Soobader, Subramanian, & Carson, 2002; Reijneveld, Verheij, & de Bakker, 2000) et en fonction de la mesure des facteurs contextuels (Chaix & Chauvin, 2003; Chaix, Guilbert, & Chauvin, 2004; Chuang, Cubbin, Ahn, & Winkleby, 2005; van Lenthe, Brug, & Mackenbach, 2005). Ces résultats illustrent que différentes échelles ou unités spatiales peuvent être pertinentes pour différentes expositions écologiques et différents indicateurs de santé (Daniel, Moore, & Kestens, 2008; Diez Roux, 2001). Bien que certains auteurs aient développé de nouvelles approches pour conceptualiser et opérationnaliser le milieu (Cockings & Martin, 2005; Frohlich, Potvin, Chabot, & Corin, 2002; Haynes, Daras, Reading, & Jones, 2007; Popay, Thomas, Williams, Bennett, Gattrell, & Bostock, 2003), leur application demeure limitée par rapport aux études qui opérationnalisent le milieu à travers les unités spatiales administratives. Dans le cadre de la recension des études multiniveaux (Riva et al., 2007), l'on constate que près de 90% des 86 études répertoriées ont employé de telles unités géographiques, dont les plus fréquentes sont les secteurs de recensement, les districts électoraux (dits « *wards* »), les territoires de codes postaux et des quartiers définis par une administration municipale.

L'utilité première de ces unités spatiales de « convenance » réside dans la facilité de les lier à des bases de données produites par les recensements ou autres enquêtes populationnelles afin de caractériser les populations qui vivent dans ces milieux.

[4] Bien entendu, ces résultats doivent être interprétés sous réserve, en raison de biais de publication possible, c.-à-d. la publication d'études rapportant des effets de milieux sur la santé qui sont statistiquement significatifs.

Aussi, puisque ces unités spatiales sont créées pour être homogènes selon des dimensions socioéconomiques, notamment les secteurs de recensement[5], elles apparaissent appropriées pour opérationnaliser ces dimensions (Ross, Tremblay, & Graham, 2004). En contrepartie, les unités spatiales de « convenance » peuvent s'avérer hétérogènes en lien avec d'autres facteurs contextuels que l'on veut étudier en relation avec la santé, tels la convivialité de l'environnement pour la pratique d'activité physique, l'accès aux services de santé, etc.; ces facteurs peuvent opérer à des échelles spatiales différentes de celles auxquelles opèrent les facteurs socioéconomiques.

Pour comprendre la signification du milieu pour la santé, il importe d'opérationnaliser des unités spatiales appropriées pour étudier les déterminants sociaux et environnementaux de la santé, c.-à-d. des unités cohérentes avec les objectifs spécifiques d'une étude soit au niveau des indicateurs de santé, des expositions ou de l'association entre exposition et santé (Chaix et al., 2006; Cummins et al., 2005a; Galea & Ahern, 2006; Gauvin et al., 2007). Qui plus est, la définition opérationnelle du milieu soulève des enjeux conceptuels et méthodologiques pouvant potentiellement limiter les validités de construit et interne des devis de recherche.

Enjeux de validité

La validité de construit consiste à déterminer si l'instrument de mesure opérationnalise le concept d'intérêt. Dans la recherche sur les effets de milieux sur la santé, la validité de construit est fonction de 1) la délimitation des unités spatiales, c.-à-d. si les unités spatiales sont appropriées pour étudier l'association entre une exposition et un indicateur de santé donné, et 2) la mesure de l'exposition, c.-à-d. si les données permettent l'opérationnalisation appropriée des variables d'exposition. L'absence de correspondance entre la distribution spatiale des variables d'exposition

[5] Cela peut ne pas être le cas pour d'autres unités administratives, comme les régions de tri d'acheminement (territoires de codes postaux canadiens) qui sont créées pour des raisons d'acheminement de courrier, et qui peuvent être hétérogènes au niveau de la composition de la population.

et la configuration des unités spatiales utilisées pour opérationnaliser les milieux peut donner lieu à des biais de mesure (ou de classification) des expositions et d'estimation des associations, et donc limiter la validité interne du devis de recherche (Osypuk & Galea, 2007).

La validité interne réfère aux caractéristiques du devis qui permettent d'inférer des associations non biaisées, donc d'obtenir une estimation fiable des effets de milieux sur la santé. Une mauvaise définition du milieu peut diminuer la validité interne d'un devis en raison d'erreurs de mesure ou de classification (non-différentielle) des expositions. Une des caractéristiques des variables spatiales est une tendance à l'autocorrélation, c.-à-d. que les milieux partagent des facteurs contextuels similaires en fonction de leur proximité dans l'espace (Cliff & Ord, 1973). En utilisant des unités spatiales qui ne tiennent pas compte de l'autocorrélation spatiale des expositions, on adopte implicitement la prémisse que l'influence sur la santé des facteurs contextuels d'un milieu est différente et s'opère indépendamment des conditions des milieux avoisinants (Coulton et al., 2001; Cummins, Curtis, Diez-Roux, & Macintyre, 2007; Diez-Roux, 2000; O'Campo, 2003; Pickett & Pearl, 2001; Riva et al., 2007; Subramanian et al., 2003). Ainsi, par l'autocorrélation spatiale des variables d'exposition, l'influence positive ou négative du milieu sur la santé peut être façonnée par le contexte des milieux voisins (Cummins et al., 2007).

De plus, une des conditions préalables pour détecter des effets de milieux, est une variation des expositions (Rothman & Greenland, 1998). Or, il se peut que cette variation soit « lissée » par la définition même du milieu. Par exemple, si des unités spatiales sont caractérisées par des populations qui sont à la fois plus nanties et d'autres qui le sont moins, la valeur « moyenne » du statut socioéconomique d'un milieu ne correspondra pas au « vrai » statut socioéconomique du milieu qui lui est hétérogène. L'homogénéité interne d'un milieu en termes d'exposition est donc requise pour minimiser les erreurs de mesure. De même, pour formuler avec justesse des inférences relativement aux effets de milieux, les différences entre les milieux doivent être maximisées : si les données sont agrégées dans des milieux contigus mais hétérogènes, l'estimation de la variation des expositions et des indicateurs de santé et leur association peuvent s'avérer erronée (*misclassification* et *misestimation* respectivement). D'ailleurs, plus les milieux sont homogènes, plus la puissance statistique se voit augmenter. Cette puissance est liée à la plausibilité dē détecter des effets. En ce sens, des effets de milieux plus importants ont été observés lorsque les unités spatiales sont plus homogènes (Haynes et al., 2007; Haynes & Gale, 1999).

L'opérationnalisation des unités spatiales est d'autant plus importante lorsqu'est considérée la problématique des unités géographiques modifiables, soit la sensibilité des résultats d'analyses à la définition opérationnelle de l'unité spatiale à laquelle les données sont agrégées (Openshaw, 1984; Openshaw & Taylor, 1979). En autres mots, les effets de milieux peuvent être observés à certaines échelles, aux échelles auxquelles les données sont colligées et agrégées, et être absents ou varier s'ils sont mesurés à d'autres échelles (Cockings & Martin, 2005; Franzini & Spears, 2003; Krieger et al., 2002; Oliver & Hayes, 2007; Reijneveld et al., 2000). Imposer des unités spatiales administratives, et possiblement arbitraires, sur des variables d'exposition autocorrélées dans l'espace peut conduire à la délimitation de patrons spatiaux artificiels ainsi qu'à une classification erronée des expositions (*misclassification*).

Dans cette optique, la création d'unités spatiales qui maximisent l'homogénéité des expositions appert comme une des stratégies pour contrer la problématique des unités géographiques modifiables (Cockings & Martin, 2005; Haynes et al., 2007; Haynes, Jones, Reading, Daras, & Emond, 2008; Openshaw, 1977; Openshaw, 1984; Openshaw, 1996). Pour ce faire, diverses approches peuvent être appliquées dont le groupement discursif, voire manuel, d'unités spatiales de base (*building blocks*) (Reading, Langford, Haynes, & Lovett, 1999; The Project on Human Development in Chicago Neighborhoods http://www.icpsr.umich.edu/Phdcn), la combinaison de méthodes statistiques et d'analyses spatiales (Law, Wilson, Eyles, Elliott, Jerrett, Moffat et al., 2005; Tatalovich, Wilson, Milam, Jerrett, & McConnell, 2006) ou l'utilisation de logiciels spécialisés de zonage automatique (Cockings & Martin, 2005; Haynes et al., 2007; Haynes et al., 2008; Martin, Nolan, & Tranmer, 2001; Nakaya, 2000; Openshaw, 1996; Openshaw & Rao, 1995).

Établir la validité des unités spatiales d'analyse appert comme une considération méthodologique majeure dans l'étude des effets de milieux sur la santé. Toutefois, elle a jusqu'à maintenant été peu abordée. Par ailleurs, la stratégie qui consiste à créer des unités spatiales qui maximisent l'homogénéité des expositions semble prometteuse pour surmonter le défi méthodologique de l'opérationnalisation des milieux (Diez-Roux, 2001; Galea, 2006; Gauvin et al., 2007; Chaix et al., 2006; Cummins et al., 2005).

La marche, un indicateur de santé important

Bien que plusieurs phénomènes de santé semblent épouser une distribution spatiale particulière, peu d'états ou d'indicateurs de santé se prêtent aussi bien à l'étude des effets de milieux que la marche. En effet, parmi tous les milieux de vie dans lesquels la marche peut être pratiquée, il semble que celle-ci soit plus souvent pratiquée aux environs du domicile (Humpel, Owen, Leslie, Marshall, Bauman, & Sallis, 2004). Aussi, la marche peut être pratiquée pour différents motifs, utilitaires ou récréatifs (Ewing, Schmid, Killingsworth, Zlot, & Raudenbush, 2003; Gauvin, Riva, Barnett, Richard, Craig, Spivock et al., 2008; Giles-Corti, Broomhall, Knuiman, & al., 2005; C. Lee & Vernez Moudon, 2006; Saelens, Sallis, & Frank, 2003; Transportation Research Board and Institute of Medicine of the National Academies [TRB & IOM], 2005), qui sont influencés par différentes opportunités présentes dans les milieux (Ewing et al., 2003; Giles-Corti et al., 2005; Pikora, Giles-Corti, Knuiman, Bull, Jamrozik, & Donovan, 2006). Les paragraphes suivants proposent un aperçu de l'importance de la pratique régulière de la marche pour la santé publique et des déterminants environnementaux qui l'influencent.

Au Québec, en 2003, 37% de la population adulte (18 ans et plus) était suffisamment active durant les loisirs, 37% était moyennement ou peu active, et 26% était inactive (Nolin & Hamel, 2005). Près de la moitié des adultes (45%) affirme utiliser la marche comme moyen de transport pour une durée d'une heure et plus par semaine (14% marchant plus de six heures par semaine). De plus, il existe des inégalités sociales considérables dans la pratique d'activités physiques; les femmes, les personnes plus âgées et celles ayant une faible scolarité et un faible revenu sont moins actives durant les loisirs (Center for Disease Control and Prevention, 2003; Nolin & Hamel, 2005).

Dans la recherche portant sur les effets de milieux sur la santé, l'identification des déterminants environnementaux de la marche fait l'objet d'un intérêt croissant. Pratiquée de façon régulière, au moins 30 minutes par jour cinq jours et plus par semaine, la marche confère des bénéfices pour la santé (Dunn, Marcus, Kampert, Garcia, Kohl, & Blair, 1999; Hakim, Petrovitch, Burchfiel, Ross, Rodriguez, White et al., 1998; Hu, Siga, Rich-Edwards, Colditz, Solomon, Willet et al., 1999; I. Lee & Paffenbarger, 2000; Pate, Pratt, Blair, Haskell, Macera, Bouchard et al., 1995; Stampfer, Hu, Manson, Rimm, & Willet, 2000; United States Department of Health and Human Services [USDHHS], 1996). Lorsque des individus sédentaires deviennent modérément actifs, les bénéfices pour la santé augmentent plus rapidement que pour les individus modérément actifs devenant très actifs (USDHHS, 1996). De plus, la marche est une activité physique accessible pour plusieurs (ne requiert pas d'habilité ou agilité particulière parmi les populations n'ayant pas de problème de mobilité) et sa pratique impose peu de contraintes (énergétique, temporelle, monétaire). Pour ces raisons, promouvoir la pratique régulière de la marche représente une composante utile d'une stratégie de santé publique pour combattre la sédentarité (TRB & IOM, 2005; USDHHS, 1996).

L'environnement bâti des milieux résidentiels est perçu comme un déterminant important de la pratique de la marche (Frank et al., 2005; Saelens et al., 2003; TRB & IOM, 2005). Selon certains auteurs, l'environnement bâti comprend au moins trois composantes : la configuration de l'occupation du sol, le système de transport et le design urbain qui, ensemble, confèrent des opportunités pour la marche (TRB & IOM, 2005). La configuration de l'occupation du sol réfère à la distribution des activités humaines dans l'espace et est liée aux concepts de densité (par exemple, de population, d'emplois), de diversité des occupations du sol et d'accessibilité aux services. La notion de distance (proximité) des activités de la vie quotidienne par rapport au lieu de résidence est fondamentale à ces concepts (Handy, Boarnet, Ewing, & Killingsworth, 2002). Le système de transport correspond aux infrastructures physiques et services qui relient les différentes activités. Le design urbain réfère aux attributs physiques et fonctionnels de l'environnement bâti, tels le style architectural des édifices, le design des rues et des espaces publics, et inclut des éléments comme les trottoirs, la connectivité du réseau de rues et l'esthétique. Le design urbain est relié de près à la configuration de l'occupation du sol et au système de transport. Ces trois composantes de l'environnement bâti peuvent à la fois faciliter et contraindre la pratique de la marche et leur influence peut varier selon les raisons (utilitaire vs récréative) qui la motivent (Humpel et al., 2004; Owen, Humpel, Leslie, Bauman, & Sallis, 2004).

Bien que les mesures dites objectives (par opposition aux perceptions que les gens ont de leur milieu résidentiel) des caractéristiques de l'environnement bâti varient selon les études, l'état actuel des connaissances indique qu'une plus grande densité de population, une occupation du sol mixte et diversifiée, un plus grand nombre et variété de destinations, la connectivité du réseau de rues (réseau en damier) et la présence de trottoirs sont associés à des fréquences et durées plus importantes des activités de marche (Cervero, 1996; Cerin, Leslie, du Toit, Owen, & Frank, 2007; De Bourdeaudhuij, Sallis, & Saelens, 2003; Ewing et al., 2003; Fisher, Li, & Michael, 2004; Frank & Pivo, 1995; Frank et al., 2005; Gauvin et al., 2008; Giles-Corti et al., 2005; Giles-Corti & Donovan, 2002a, 2003; Humpel et al., 2004; C. Lee & Vernez Moudon, 2006; Leslie, Coffee, Frank, Owen, Bauman, & Hugo, 2007; Li, Fisher, Brownson, & Bosworth, 2005; Lovasi, Vernez-Moudon, Pearson, Hurvitz, Larson, Siscovick, et al., 2008; Owen et al., 2004; Pikora et al., 2006; Saelens et al., 2003; van Lenthe et al., 2005).

Pour atteindre les recommandations d'activités physiques, il est possible de cumuler, au cours de la journée, des périodes d'activités physiques d'au moins 10 minutes, totalisant un minimum de 30 minutes (TRB & IOM, 2005; USDHHS, 1996). Ainsi, plutôt que d'être confinée à une plage horaire précise, la pratique d'activités physiques peut être intégrée aux activités quotidiennes conférant ainsi importance au concept de « vie active ». De la sorte, les niveaux souhaitables d'activités physiques peuvent être atteints à travers une multitude d'activités comme la pratique de la marche ou du vélo pour se déplacer, pour l'exercice, ou pour les loisirs, vaquer à des tâches à l'intérieur et à l'extérieur du domicile, emprunter les escaliers plutôt que l'ascenseur, utiliser des installations récréatives, etc. (Active Living Research http://www.activelivingresearch.org). Le concept de « vie active » met de l'avant l'idée de rendre ubiquitaires les opportunités pour être actif de telle sorte que l'activité physique soit intégrée aux routines quotidiennes. Dans cette optique, les caractéristiques de l'environnement bâti peuvent être perçues comme ayant le potentiel de réguler la propension à une vie active des individus et des populations (Gauvin, Richard, Craig, Spivock, Riva, Forster et al., 2005).

Pour mieux informer les décideurs responsables de la conception et de l'implantation des politiques de santé publique et des initiatives de promotion de la santé visant à créer des environnements qui favorisent un mode de vie actif, il importe de comprendre le rôle des milieux résidentiels dans la promotion de la marche. Cela requiert l'opérationnalisation d'unités spatiales homogènes au niveau des expositions d'intérêt.

QUESTION ET OBJECTIFS DE RECHERCHE

Comme il a été énoncé en introduction, cette thèse porte sur la question de recherche suivante : *Quel est l'apport du développement et de l'application d'une approche novatrice pour créer des unités spatiales pour la compréhension des effets de milieux sur la santé ?* Pour répondre à cette question, deux objectifs spécifiques ont été poursuivis :

1) Créer des unités spatiales optimales homogènes au niveau du potentiel de vie active des milieux, opérationnalisé par la densité de la population, la mixité de l'occupation du sol et l'accessibilité géographique[6] aux services de proximité; et quantifier et comparer les degrés d'homogénéité des nouvelles unités et des secteurs de recensement au niveau du potentiel de vie active des milieux.

2) Examiner l'influence des milieux sur les comportements de marche d'adultes montréalais de 45 ans et plus, en comparant les résultats obtenus lorsque le milieu est opérationnalisé par les nouvelles unités spatiales plutôt que par les secteurs de recensement

Ce projet est pertinent pour deux raisons. D'abord, comme le signalent plusieurs auteurs (Coulton et al., 2001; Diez-Roux, 2000; Diez Roux, 2001; Gauvin et al., 2007; O'Campo, 2003; Pickett & Pearl, 2001; Riva et al., 2007; Subramanian, 2004; Subramanian et al., 2003), l'opérationnalisation des milieux dans la première génération d'études a, pour la plupart, été réalisée selon des méthodes de convenance, dictées par la disponibilité des unités spatiales définies pour des motifs autres qu'étiologiques. Ce projet de doctorat propose une méthode novatrice pour créer des unités spatiales ayant une signification étiologique. Dans cette optique, il a été suggéré que l'opérationnalisation d'unités spatiales soit en cohérence avec les

[6] L'accessibilité géographique réfère à la facilité avec laquelle les résidents d'un milieu peuvent atteindre les services et équipements collectifs (Hewko et al., 2002). Pour alléger le texte, le terme accessibilité sera parfois utilisé seul.

objectifs spécifiques de l'étude (Chaix et al., 2006; Cummins et al., 2005a; Diez Roux, 2001; Galea & Ahern, 2006; Gauvin et al., 2007). En délimitant des unités spatiales en fonction de la distribution spatiale du potentiel de vie active des milieux associé aux comportements de marche, c'est selon cette approche « spécifique » que la création de zones a été développée et appliquée.

Contexte de la thèse

Les données individuelles analysées proviennent du projet de recherche « *Physical/environmental structures and walking behaviour in residential living spaces: Exploring pathways through which place influences a specific health behaviour* » (subvention # 200203 MOP 57805) (projet MARCHE) financé par les Instituts de recherche en santé du Canada dont Lise Gauvin est la chercheure principale. Le sommaire de ce projet est présenté à l'Annexe II. L'objectif du projet MARCHE est de mieux comprendre la relation entre les environnements physiques et la marche dans des quartiers urbains et de banlieues. Le projet a été développé en deux phases : (1) observation systématique d'aspects spécifiques du potentiel de vie active de 112 quartiers résidentiels (définis par les secteurs de recensement), soit la densité des destinations, la convivialité et la sécurité (Gauvin et al., 2005); (2) recrutement de 2923 personnes de 45 ans et plus et vivant dans les 112 secteurs de recensement pour participer à une entrevue téléphonique visant à connaître leurs pratiques de marche, leurs autres habitudes de vie et leurs caractéristiques sociodémographiques. Les associations entre la densité des destinations, la convivialité et la sécurité des quartiers et la pratique régulière de la marche (au moins 30 minutes par jour, trois, cinq ou sept jours par semaine) ont été analysées à l'aide de modèles statistiques multiniveaux (Gauvin et al., 2008).

Par rapport au projet MARCHE, cette thèse a une portée unique. D'abord, les objectifs et la finalité sont distincts. Le but de cette thèse est de déterminer l'apport du développement et de l'application d'une approche novatrice pour créer des unités spatiales pour la compréhension des effets de milieux sur la santé. À cette fin, la marche et ses déterminants environnementaux sont utilisés à titre d'exemple. Ensuite, le potentiel de vie active a été défini par trois mesures objectives provenant de bases de données spatiales, soit la densité de la population, la mixité de l'occupation du sol et l'accessibilité aux services; ces mesures sont indépendantes des données observationnelles du projet MARCHE. Finalement, les données ont été traitées à l'aide de différentes méthodes d'analyses spatiales (analyse de réseaux) et statistiques (classification statistique, modèles multiniveaux avec classification croisée). Dès lors, il est clair que les données provenant du projet de recherche MARCHE ont été explorées de façons différentes et novatrices.

Le protocole de recherche du projet MARCHE a été approuvé par le Comité d'éthique sur la recherche chez les êtres humains de la Faculté de médecine de l'Université de Montréal à l'été 2004 (Annexe III). L'accès aux données spatiales est public ou régi par l'entremise d'ententes interuniversitaires. Les résultats ne divulguent pas la localisation spatiale des individus ni des secteurs de recensement dans lesquels les participants ont été recrutés.

ABSTRACT

Background

In health and place research, definitions of areas, area characteristics, and health outcomes should ideally be coherent with one another. Yet current approaches for delimiting areas mostly rely on spatial units "of convenience" such as census tracts. These areas may be homogeneous along socioeconomic conditions but heterogeneous along other environmental characteristics. This heterogeneity can lead to biased measurement of environment characteristics and misestimation of area effects on health. The objective of this study was to assess the soundness of census tracts as units of analysis for measuring the active living potential of environments, hypothesised to be associated with walking.

Results

Starting with data at the smallest census area level available, zones homogeneous along three indicators of active living potential were designed: population density, land use mix, and accessibility to services. Delimitation of zones ensued from statistical clustering of the smallest areas into seven clusters or "types of environment". Mapping of clusters into a GIS led to the delineation of 898 zones characterised by one of seven types of environment, corresponding to different levels of active living potential. Homogeneity of census tracts along indicators of active living potential varied. A greater proportion (83%) of variation in accessibility to services was attributable to differences between census tracts suggesting within-tract homogeneity along this variable. However, census tracts were heterogeneous with respect to population density and land use mix where a greater proportion of the variation was attributable to within-tract differences. About 55% of tracts were characterised by a combination of three or more "types of environment" suggesting substantial within-tract heterogeneity in the active living potential of environments.

Conclusion

Soundness of census tracts for measuring active living potential may be limited. Measuring active living potential with error may lead to misestimation of associations with walking, therefore limiting the correctness of inference about area effects on walking. Future studies should aim to determine homogeneity of spatial units "of convenience" along environment characteristics of interest prior to examining their association with health. Further evidence is needed to assess the extent of this methodological issue with other indicators of environment context relevant to other health indicators.

INTRODUCTION

Residential areas are proximal to everyday life and are therefore likely to influence health of local populations through the possibility they provide for leading healthy lives [1, 2]. An accumulating body of research shows evidence for variation in health across residential areas and the significance of area context for explaining this variation, independently of the characteristics of individuals [3-5].

Different scales, or spatial units, may be relevant to specific contextual conditions and to specific heath outcomes [6, 7], as illustrated by studies reporting varying strength and magnitude of area effects on health according to the operational definition of areas [8-15] or to contextual conditions [16-19]. Nonetheless current approaches for delimiting areas mostly rely on spatial units "of convenience" such as census tracts, boroughs, or wards [3, 5]. These spatial units are certainly useful because they can easily be linked to data from censuses and other surveys that can be used for measuring contextual conditions. Also, they are often designed to be homogeneous along socioeconomic conditions of populations, thus being appropriate spatial units to operationalise the socioeconomic context of areas [20] (yet this may not hold for other administrative units, e.g. postal code areas which are design for postal delivery purposes and may be very heterogeneous in terms of population composition). However it is to be considered that through time, the composition of the units may change leading to modification of the socioeconomic conditions which may become more heterogeneous.

Yet, other contextual dimensions relevant for health may not be optimally defined within administrative spatial units. For example, conduciveness of areas to physical activity or geographic accessibility to health services may operate on different scales than socioeconomic factors. Operationalising relevant spatial units for studying area effects on health remains a conceptual and methodological challenge [4, 5, 7, 21-26] giving rise to issues of validity and soundness of areal units as units of analysis [27].

Operationalising small areas: issues of validity and soundness units of analysis

Construct validity refers to whether or not the measurement instrument operationalises the concept of interest. In area effects on health research, construct validity is a matter of establishing 1) the soundness of units of analysis, i.e., whether or not area boundaries are aetiologically meaningful for studying the association between area characteristics and a given health indicator, and 2) whether or not data constitute appropriate operationalisations of exposure variables, i.e. the characteristics of areas [27]. Ideally, definitions of areas, the characteristics of these areas, and the health outcome(s) being studied should be coherent with one another [7].

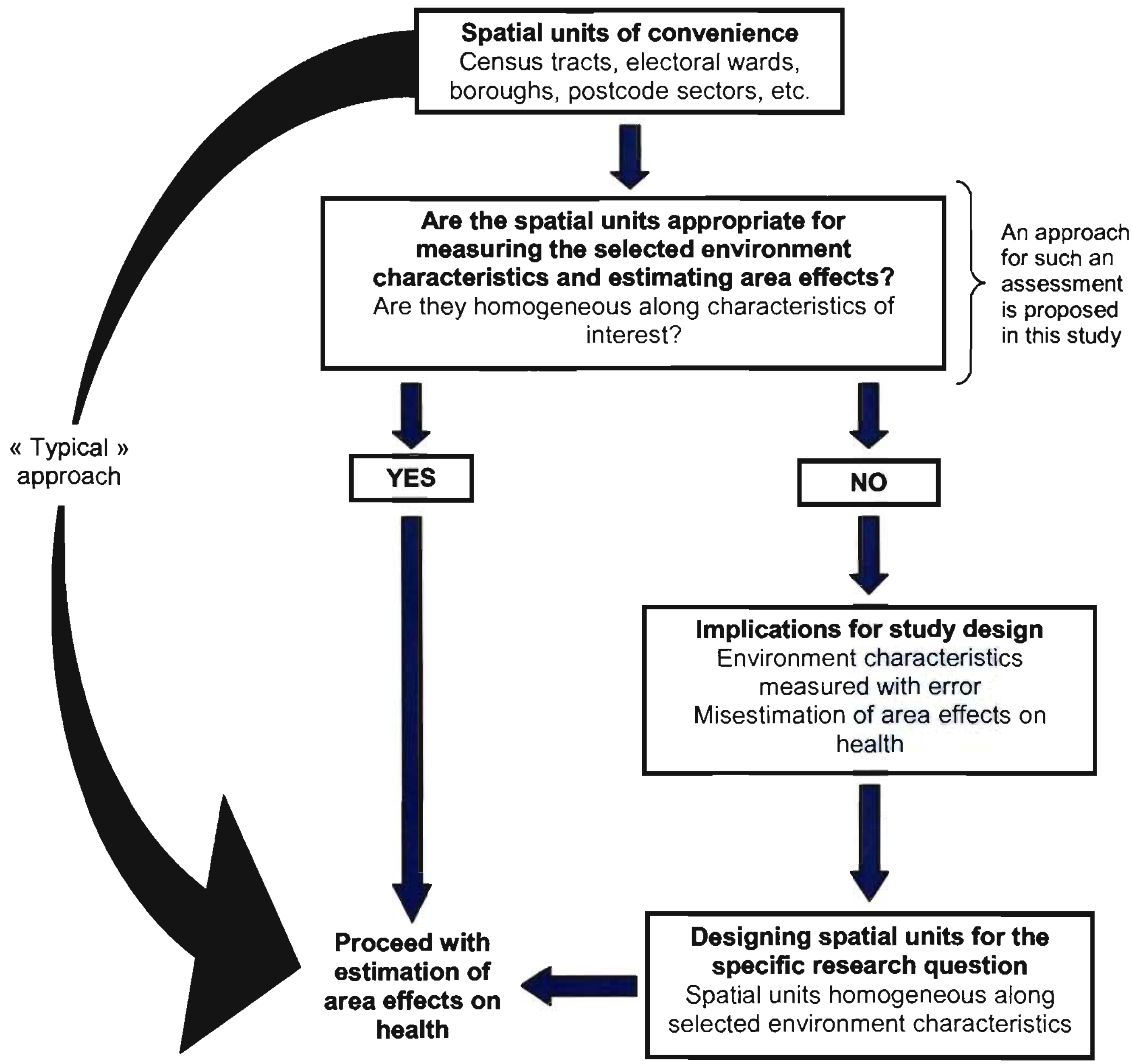
Measures of area characteristics derived from census and other surveys, e.g. socioeconomic position, although easily accessible, provide only partial information on the context of areas and may in fact be endogenous to the composition of the areas as they are determined by individual characteristics of residents [28]. Collecting and measuring "true" or "integral" area data, i.e. data only measurable at the area level through such procedures as ecometrics and spatial analysis has been underscored as critical for measuring unbiased area-level variables [2, 7, 28, 29]. Likewise, defining aetiologically meaningful areas in coherence with the specific purposes of the study, either in terms of health outcomes, characteristics of environment, or associations between the two [23, 28, 30, 31] is important for understanding the significance of residential areas for health. Measurement errors can result if the spatial patterning of environmental characteristics does not correspond to the spatial units chosen for operationalising areas and their context [27].

Defining relevant geographic areas becomes salient in light of the modifiable areal unit problem, i.e. the fact that analytical results are sensitive to the definition of spatial units at which data are aggregated [32, 33]. In other words, area effects may be observed only at certain scales, i.e. scales at which data are collected and aggregated and may vary or be absent when observed at other scales. Imposing arbitrary spatial units on a continuous spatial process, e.g. characteristics of environments, may lead to the delineation of artificial spatial patterns. In such cases, environment characteristics may be measured with error. As a result the internal validity of the study, i.e. whether or not observed associations are unbiased, may be threatened.

In addition, as per spatial autocorrelation, areas will share similar contextual conditions as a function of their proximity in space [34]. By using spatial units of convenience, it is assumed that contextual conditions within one area are different and influence health independently of conditions in neighbouring areas [4, 5, 21, 22, 24-26, 35], when in fact these conditions are clustered in space. Furthermore, for any area effects to be detected there must be variation in the exposure being studied [36]. Yet the variation of environment characteristics may be smoothed out by the definition of area units used to measure them. For example, if spatial units encompass environments that are both conducive to walking and others that are less so, averaging values of conduciveness over census tracts could potentially lead to mismeasurment of exposures. Within area homogeneity along the contextual conditions under examination is thus required for minimising measurement error. Correspondingly, for inferring about area effects on health, between area differences must be maximised: if data are collected in contiguous and heterogeneous areas, variations in both characteristics of environments and health outcomes, and their association, may be misestimated. As area effects on health have been observed to be stronger in more homogeneous areas [37, 38], homogeneity of areas may thus influence the estimation of area effects and therefore the validity of conclusion.

In Figure 1, we propose a template that could be useful for establishing the soundness of spatial units "of convenience" to operationally define areas for specific research questions. For example, the template could be used to guide the decision as to whether or not census tracts are the most appropriate spatial units of analysis for

Figure 1. Template for deciding upon the soundness of spatial units "of convenience" to operationally define small area units of analysis



measuring associations between area-level socioeconomic position (SEP) and obesity. That is, if they allow for measuring indicators of SEP without bias (and ultimately for estimating non-biased association with health outcomes) by showing homogeneity in the distribution of indicators and optimising their spatial patterning. In the methods sections, we propose an approach for achieving this end. Intuitively, it can be expected that census tracts are appropriate units for undertaking such a study as they are, as mentioned above, initially designed to be homogeneous along socioeconomic conditions. But across time, the socioeconomic composition of census tracts may change as people migrate in and out of areas, potentially introducing heterogeneity in the socioeconomic make-up of the area. This could result in a "dilution" of the true level of deprivation. Averaging indicators of SEP over census tracts thus may mask "pockets" of poverty. The exercise of establishing the soundness of census tracts as units of analysis would be important here, as it would allow to measure with less error indicators of SEP and their association with health outcomes. In multilevel studies, mismeasurement of environment characteristics may influence the strength of the observed association between environment characteristics and health indicators [39]. As such, associations may not be detected or may be spurious, therefore limiting the precision of research findings for informing public health and public policy actions to tackle social and geographical inequalities in health.

Establishing the soundness of spatial units of analysis chosen for operationalising area boundaries and measuring area context is an important methodological consideration, but it is often overlooked. Alternatively, designing spatial units of analysis maximising homogeneity of selected environment characteristics may prove to be a viable strategy for advancing the understanding of processes linking place to health [7].

OBJECTIVES

The aim of this investigation is to assess the soundness of census tracts as units of analysis for studying associations between a specific exposure and a specific health outcome, namely the active living potential of residential environments and walking behaviours. Active living potential refers to the conditions of areas that encourage the likelihood of integrating physical activity into daily routines [29]. Census tracts were selected as spatial units "of convenience" because of extensive use of this spatial unit of analysis in current research on health and place [4, 5]. In Canada, census tracts are small and relatively stable geographic areas with populations ranging in size between 2500 and 8000 inhabitants; at the time of their creation, census tracts were homogeneous in terms of socioeconomic characteristics, e.g. economic status and social living conditions [40].

To establish the soundness of census tracts as a unit of analysis, we developed and tested a comprehensible method for designing optimal and homogeneous spatial units espousing the spatial distribution of selected environment characteristics linked to the concept of density of destinations that is the physical and social characteristics of residential areas related to land use pattern [29]. Three indicators were used to operationalise the construct of active living potential: population density, land use mix, and geographic accessibility to proximity services. The specific objectives of the study are to examine whether or not: 1) census tracts are homogeneous units of analysis along indicators of active living potential; 2) active living potential and socioeconomic indicators follow a similar spatial distribution; and 3) census tracts encompass smaller areas or units of analysis with different (or similar) levels of active living potential.

Active living potential was chosen because of increasing research reporting associations between this environmental construct and walking [19, 41-53], an important public health indicator [54-56]. This choice was also motivated by availability of spatial datasets allowing for the operationalisation of integral measures of land use mix and geographic accessibility to services in geographical information systems, and by the availability of individual-level data on walking behaviours (to be examined in future analyses).

METHODS

The methodology section includes two parts. First, we present criteria and methods for designing homogeneous areas (henceforth designated as "zones"). Second, we present analyses undertaken to assess the soundness of census tracts as units of analysis for measuring the active living potential of residential areas.

Designing optimal, homogeneous zones

Zone design refers to the placement of areal unit boundaries [9]. It can be achieved discursively (manually) by grouping basic spatial units into larger ones [57-59], by combining social, statistical, and spatial analysis methods [60, 61], and automatically through computationally intensive automated zoning software [9, 15, 37, 62-65].

Three criteria guided the choice of the method for zone design. First, we wanted to design zones based on the spatial distribution of environmental characteristics related to active living potential, namely population density, land use mix, and geographic accessibility to selected proximity services. We had no requirement regarding population and area sizes as zones were defined on the basis of the spatial distribution of these characteristics. Second, the method for zone design had to be optimal, i.e. to maximize variation between zones and to minimize variation within zones in the selected characteristics. In other words, the aim was to design zones that were internally homogeneous on the three indicators of active living potential, but different (heterogeneous) amongst themselves. Finally, we wanted a method that was rigorous but comprehensible and easy to implement. We opted for an approach that combined a statistical classification method, *K*-means clustering, to mapping applications in geographic information system. This three-step approach is described in greater details in the following sections.

Step 1: Measuring environment characteristics at the smallest area level

The study area is the Island of Montreal, Canada, an urban centre with 1 812 723 residents. As of January 2006, on the Island of Montreal, there are 15 municipalities, in addition to the municipality of Montreal which includes 19 boroughs [66]. The Island of Montreal is further divided into 521 census tracts and 3222 dissemination areas. Dissemination areas (DAs) were used as basic spatial units for designing zones because they are the smallest standard geographic areas for which Canadian census data are available. On the Island of Montreal, their average size is 0.15 km^2 (ranging between 364 m^2 and 18 km^2) with an average population of 562 individuals (ranging between 44 and 2138 residents). DA values for population density, land use mix, and accessibility of services were computed in a geographical information system (ArcGIS 9.2) [67].

Population density refers to number of individuals per unit area. It was computed by dividing the total number of residents of a DA by its area size (km^2) [68].

Land use mix relates to the diversity or variety of land uses within an area. It was computed using an entropy index [47, 69, 70] which measures the homogeneity or diversity of land uses within a spatial unit. The index is defined as follow:

$$E_j = -\sum_{i=1}^{n} \left[(A_{ij} / D_j) \ln (A_{ij} / D_j) \right] / \ln n \qquad \text{(Equation 1)}$$

Where A_{ij} is the surface area of land use i in dissemination area j, D_j is the surface area of dissemination area j, and n is the total number of possible land uses which in the current case corresponds to 16, the number of different land uses characterising the Island of Montreal [71]. The index values range between 0 and 1, where 1 corresponds to a highly mixed area, and 0 to a homogeneous area, that is an area characterised by only one land use (e.g. low density housing). This index has been used in many studies to measure land use mix [47, 72].

Geographic accessibility to proximity services refers to geographic distance to or from destinations, here to supermarkets, pharmacies, banks, and libraries. These services were selected because they are most likely to be used on a regular basis, conveying the idea of proximity services potentially accessible through walking. There are many measures of geographic accessibility [73, 74]. In this study, geographic accessibility was defined in terms of the number of the selected services within an area, conferring the notion of the offer of services provided by the immediate surroundings. Supermarkets, pharmacies, banks, and libraries were geocoded at the parcel level [75]. In order to minimise aggregation errors [73, 76], accessibility was measured by computing distances of services located within a one kilometre (network distance) radius [77] from the centroid of census blocks (n=14 527) comprised within any one DA; the distances were than averaged and weighted by the total population of each census blocks.

Characterisation of DAs along the three indicators resulted in a sample of 3206 DAs. Measures of land use mix and accessibility to services were normally distributed; population density was normalized using a LOG10 transformation [78]. Population density was significantly and positively correlated to accessibility to services ($r=0.45$, $p<0.001$), and negatively to land use mix ($r=-0.32$, $p<0.001$). Land use mix and accessibility to services were not significantly correlated ($r=0.03$, $p>0.500$). Prior to cluster analyses, these variables were standardized to a mean of 0 and a standard deviation of 1, higher values representing greater levels of population density, land use mix, and accessibility to services.

Step 2: Classifying smallest areas into clusters, e.g. "types of environments", using K-means clustering

K-means statistical clustering techniques using SAS (version 9.1) for Windows [79] was applied to classify DAs into *k* number of optimal clusters homogeneous in terms of active living potential. In social sciences, notably in geography, *K*-means is largely employed to classify areas (e.g. geodemographics [80]). The method uses an allocation/re-allocation algorithm to optimally reassign objects, here DAs, to the nearest cluster centroid [81-83]. The goal is to maximize between cluster variations and to minimize within cluster variations. The aim of this second step was to group DAs with similar values of population density, land use mix, and accessibility of services into *k* types of environments that are internally homogeneous but different among them. These types of environments correspond to different levels of active living potential. In *K*-means, the number of clusters (*k*) must be determined at the onset of analyses; as we had no a priori for such number, we conducted analyses for k=4 to k=20.

Step 3: Mapping the clusters to create optimal and homogeneous zones

In a final step, the *k* types of environments were imported into ArcGIS 9.2 and mapped out. This lead to the delineation of *n* homogeneous zones i.e., units of analysis, characterised by one of *k* active living potential.

Statistical analyses: Assessing the soundness of census tracts as units of analysis for operationalising active living potential

The soundness of census tracts for operationalising indicators of active living potential was assessed through three series of analyses.

First, to assess the homogeneity of census tracts, variation in indicators of active living potential was estimated and decomposed between and within areas. Population density, land use mix, and accessibility of services were measured continuously at the DAs level (level 1: n=3206). In separate two-level multilevel models, DAs were nested into zones (n=898) and into census tracts (n=506 with valid population and socioeconomic data). Between-area variation in indicators of active living potential was estimated using the intraclass correlation coefficient (ICC) from unconditional (null) multilevel models using HLM software Version 6.04 [84]. The ICC indicates the proportion of variation in a dependent variable that is attributable to differences between area units. Greater ICC values indicate that variation of a variable is greater between units than within, i.e. units are different among them but internally homogeneous. Using the same analytical approach, homogeneity of zones and census tracts along indicators of socioeconomic position was assessed and compared. DA-level data on the proportion of low-income households, of people with less than high school education, and of people with a university degree were obtained from the 2001 Canadian census.

Second, analyses of variance were performed to examine the proportion of variation across zones in socioeconomic variables explained by the *k* types of environment. Indicators of SEP at the DA-level were aggregated (weighted by population) at the zone-level. These analyses were performed to examine whether or not socioeconomic and active living indicators follow a similar spatial distribution as is implicitly assumed when measured within the same area unit of analysis.

Finally, descriptive statistics were employed to assess the extent to which the spatial distribution of the different types of environment coincides with the boundaries of census tracts. These analyses were conducted to examine if census tracts encompassed environments with differing levels of active living potential. The numbers of zones straddling over one or more census tracts, and the number of types of environment encompassed within census tracts were computed. To examine whether or not the spatial distribution of more mixed or more homogeneous census tracts (i.e. the number of types of environments encompassed within census tracts) was structured in space, global values of spatial autocorrelation were computed using Moran *I* with a first-order contiguity matrix [85, 86]. Values for Moran *I* vary between -1 and 1, where negative values indicate negative spatial autocorrelation, i.e. neighbouring spatial units have different values, and positive values indicate positive spatial autocorrelation, i.e. neighbouring units have similar values. The covariance in Moran *I* is the covariance over space for neighbouring spatial units, and will not be computed unless two units are contiguous (first order); also, only one variable is considered [85], here the number of types of environments included in census tracts.

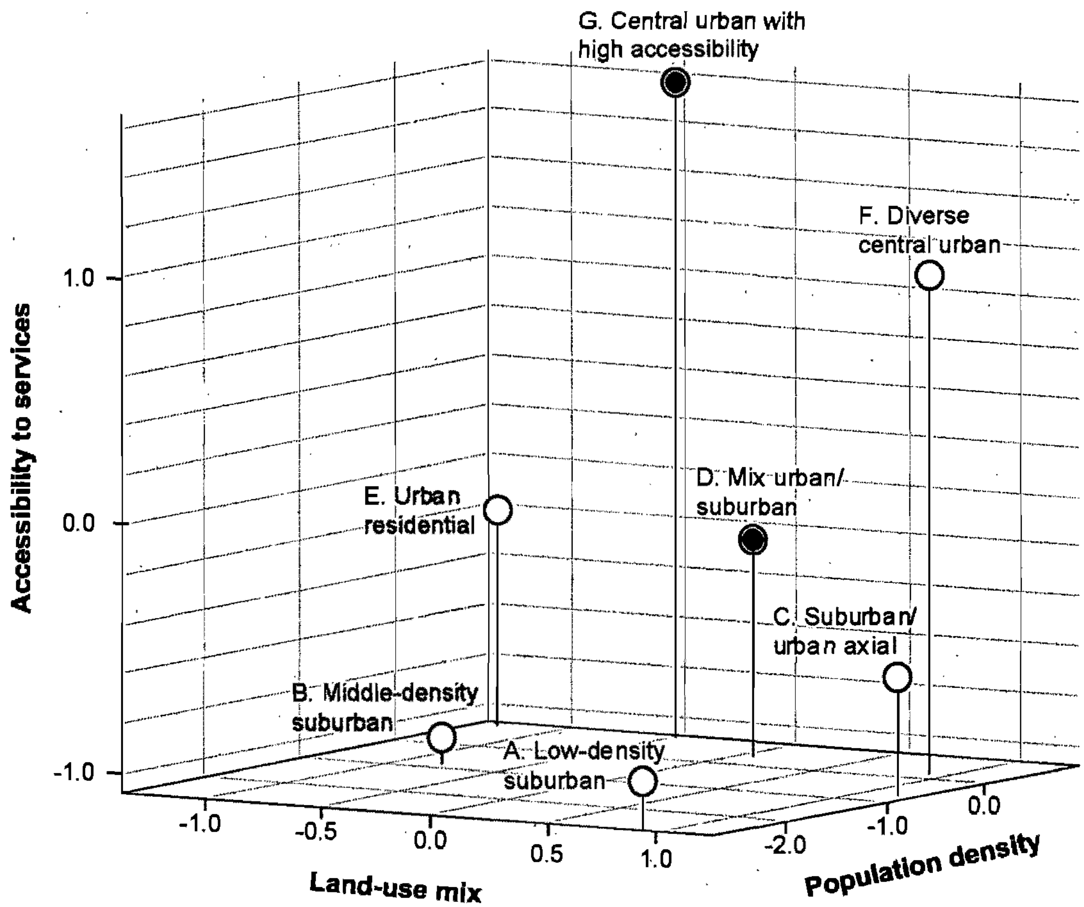
RESULTS

Description of types of environment and zones

Figure 2 illustrates results of the *K*-means clustering, which show that the 3206 DAs were optimally classified into 7 clusters or "types of environments" as indicated by peaks [87] in both the Pseudo-F statistic [88] and the Cubic clustering criterion [89]. These clusters explain 72.8% of the total variation in the three indicators of active living potential. Thus, differences among the seven clusters and similarity of DAs comprised within the same cluster, i.e. within-cluster homogeneity, were both maximized. The seven types of environments correspond to seven different levels of active living potential. They encompassed more suburban to more central urban types of environments defined by different values for population density, land use mix, and accessibility to services. The types of environments are described in Figure 2 and Figure 3.

Low-density and mid-density suburban areas are characterised by lower values of population density and accessibility to services. Diverse central urban areas and central urban areas with high accessibility are more densely populated and have greater access to services than any other types of environment. Although population

Figure 2. Statistical proximity of the seven types of environments (clusters)



Euclidian distance between the centroids of categories of exposure (statistical proximity)*

	A	B	C	D	E	F
A. Low-density suburban						
B. Middle-density suburban	2.53					
C. Suburban/urban axial	2.04	2.01				
D. Mix urban/suburban	3.20	1.63	1.54			
E. Urban residential	4.18	1.86	2.89	1.43		
F. Diverse central urban	3.40	2.76	1.78	1.44	2.56	
G. Central urban with high accessibility	4.51	3.04	3.12	1.94	1.94	1.68

* The lower the value, the closer the centroids.

density and accessibility to services follow to some extent an increasing gradient from more suburban to more urban areas, the pattern of land use mix is more complex: there are low values in urban areas and high values in suburban areas. Dissemination areas are designed to be similar in population size (among other characteristics); thus the area size required to reach the set population threshold will be larger in less densely populated areas and smaller in more urban areas. As a consequence, larger dissemination areas are more likely to encompass different land use than are smaller dissemination areas located in urban areas.

Figure 2 also presents the statistical proximity (Euclidian distance) of the centroids of clusters (cluster mean values), i.e. types of environment, in a three dimensional graph where the axes correspond to the three indicators of active living potential. With respect to their spatial distribution, the types of environment are positively correlated in space indicating that contiguous zones were characterised by similar types of environment.

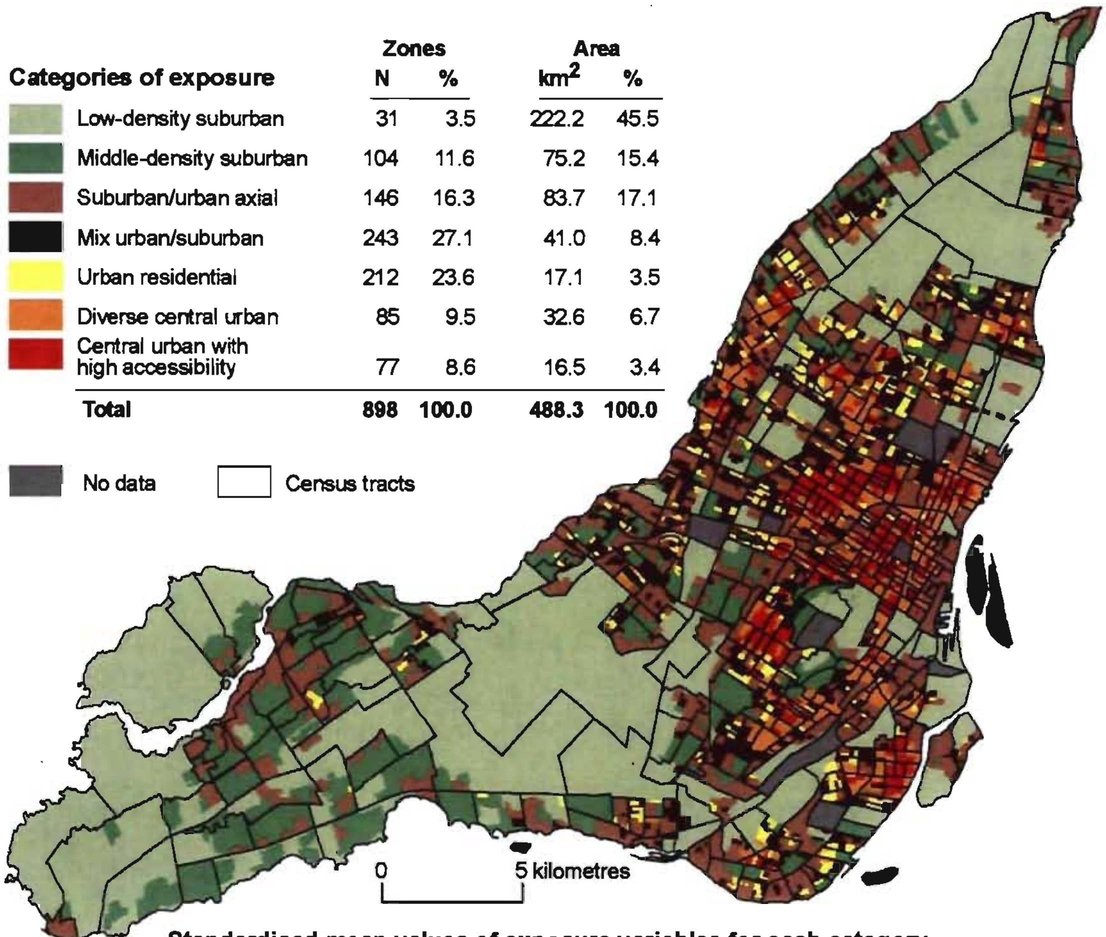
Mapping of the clusters into the GIS led to the delineation of 898 zones or units of analysis characterised by one of the seven types of environments, i.e. active living potential, as illustrated in Figure 3. Zones are significantly smaller than census tracts, an average of 0.54 km^2 (SD=3.50) compared to 0.96 km^2 (SD=1.98) ($t=-2.46$; $p<0.05$), but the range in their size is not statistically different ($F=0.68$; $p=0.409$). Zones are significantly smaller than census tracts in population size, an average of 1960 (SD=3867) residents compared to 3554 (SD=1647) ($t=-10.73$; $p<0.001$), and there is significantly greater variability in population size across zones than across census tracts ($F=11.40$; $p<0.01$). Zones characterised by more suburban contexts are on average larger and have relatively smaller population counts than urban zones.

Figure 3. Description of types of environment (clusters) and zones

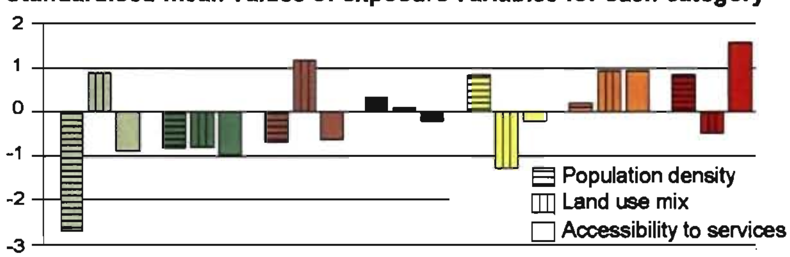
Categories of exposure	Zones N	Zones %	Area km²	Area %
Low-density suburban	31	3.5	222.2	45.5
Middle-density suburban	104	11.6	75.2	15.4
Suburban/urban axial	146	16.3	83.7	17.1
Mix urban/suburban	243	27.1	41.0	8.4
Urban residential	212	23.6	17.1	3.5
Diverse central urban	85	9.5	32.6	6.7
Central urban with high accessibility	77	8.6	16.5	3.4
Total	**898**	**100.0**	**488.3**	**100.0**

No data

Census tracts



Standardised mean values of exposure variables for each category



Description of catogories of exposure

Low-density suburban	Suburban areas characterised by lowest population density, land use diversity, and low accessibility to services
Middle-density suburban	Suburban areas characterised by low population density, land use homogeneity, and weak accessibility to services
Suburban/ urban axial	Suburban and urban areas characterised by low population density, land use diversity, and weak accessibility to services. Located close to main street axis
Mix urban/ suburban	Suburban and urban areas with average values of population density, land use mix, and accessibility to services. Located mainly in the central and eastern part of the island
Urban residential	Urban areas with highest population density, land use homogeneity, and weak accessibility to services. Located mainly in peripheral residential areas to central urban areas
Diverse central urban	Urban areas characterised average population density, land use diversity, and good accessibility to services
Central urban with high accessibility	Urban areas with high population density, land use homogeneity, and highest accessibility to services. Located mainly in central areas of the Montréal Island

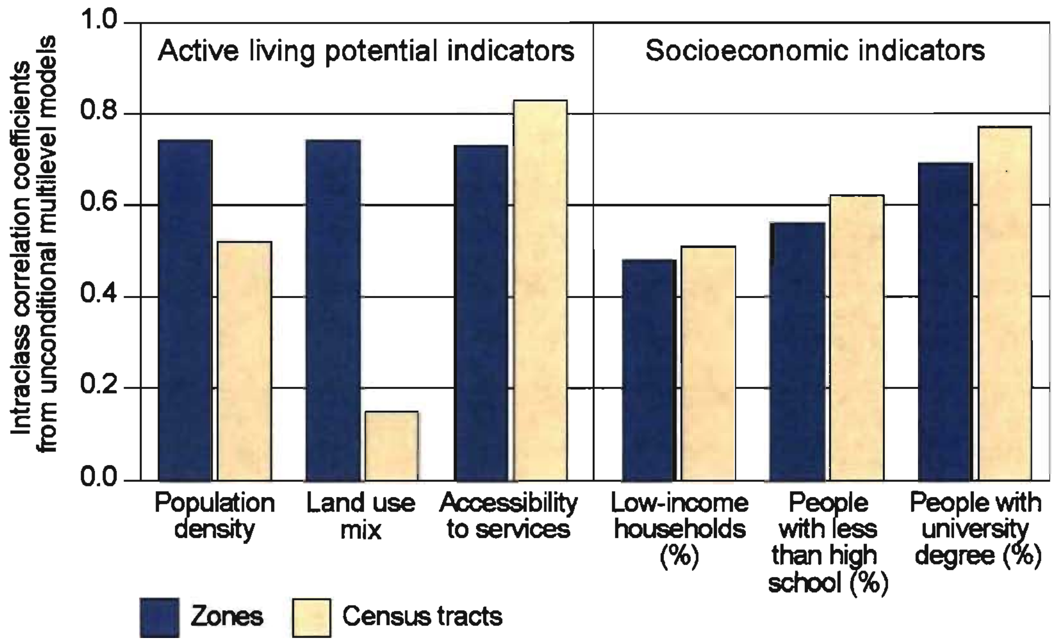
Soundness of census tracts as units of analysis for measuring active living potential

Homogeneity of census tracts along active living potential indicators

Results of homogeneity of zones and census tracts along active living indicators appear in Figure 4. The variation in indicators is not uniform across census tracts. A greater proportion (83%) of variation in accessibility to services is attributable to differences between census tracts, as indicated by a higher ICC value, suggesting within census tract homogeneity along this indicator. Yet about half of the variation in population density is between census tracts (52%), whereas there is greater variation in land use mix within census tracts (85%), indicating greater heterogeneity of tracts along these indicators. The degree of homogeneity of tracts therefore varies according to the indicator examined. For population density and land-use mix, but not for accessibility of services, variation between zones is greater than between census tracts. This shows that the method was successful in designing areas or units of analysis that were more homogeneous than census tracts along dimensions of active living potential.

The degree of homogeneity of census tracts and zones along socioeconomic indicators shows that for the selected variables, variability is larger between census tracts than between zones (Figure 4). Census tracts are relatively homogeneous areas in terms of the socioeconomic environment, especially for proportion of population with a university education.

Figure 4. Decomposition of variation in indicators of active living potential and socioeconomic position across zones and across census tracts.



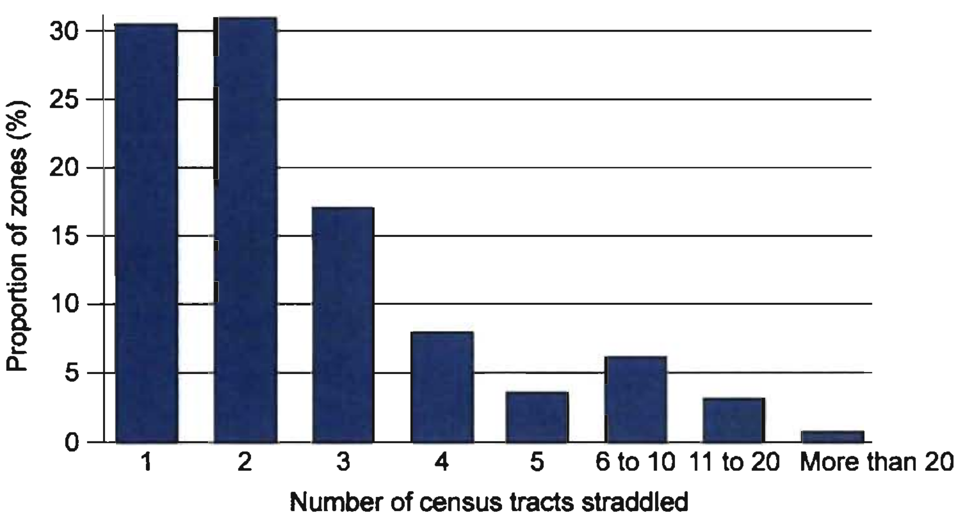
Spatial distribution of active living potential and socioeconomic indicators

Examining variation in socioeconomic indicators across zones shows that they follow a different spatial distribution than that of active living potential indicators. Results of analyses of variance (results not shown) revealed that 15.2% of the variation in the proportion of low-income households was explained by the seven types of environment whereas these proportions were 5.2% for the proportion of people with less than high school and 3.8% for the proportion of people with a university education.

Types of environments encompassed within census tracts boundaries

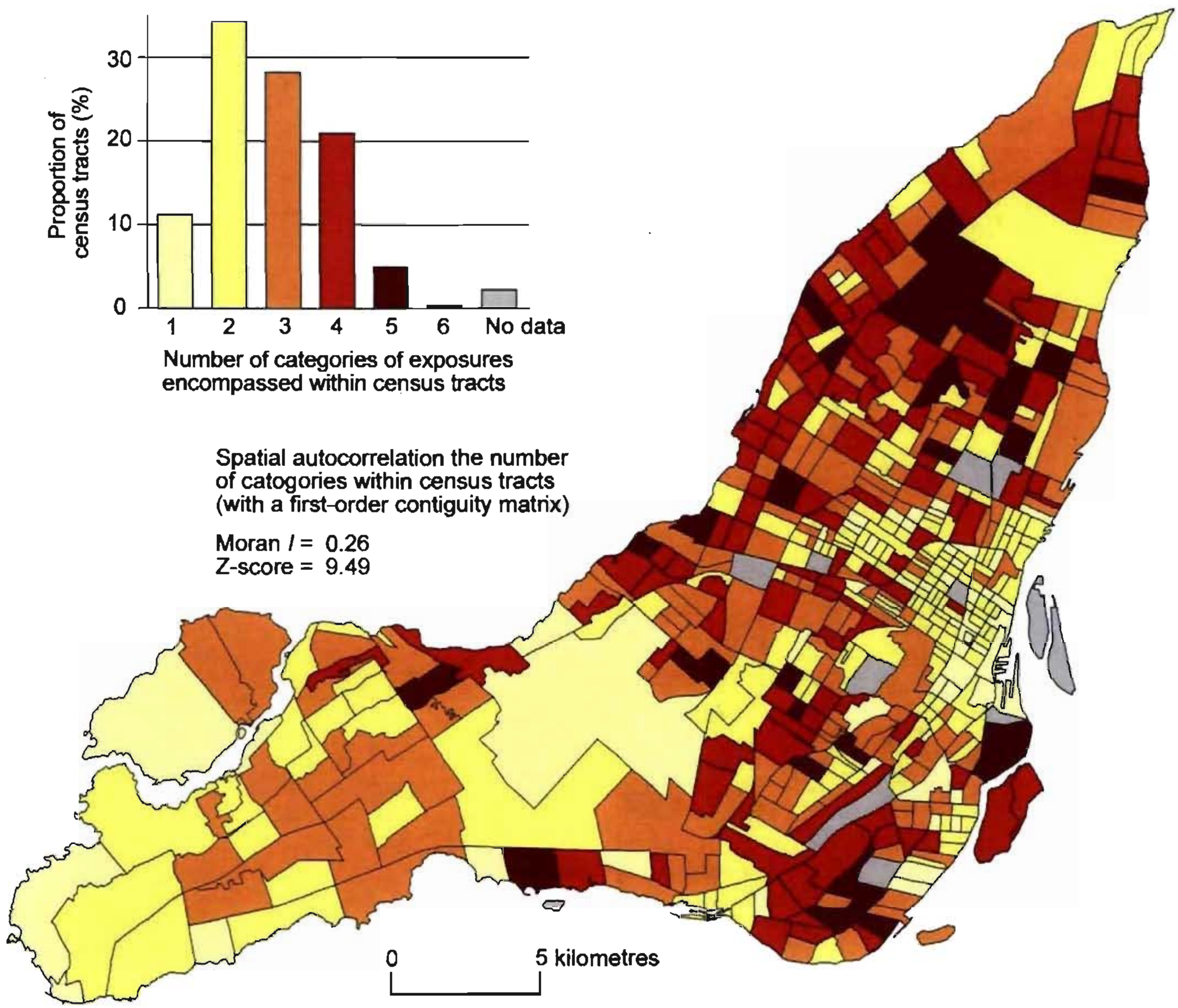
Overall, zones are not well contained within census tracts. As shown in Figure 5, only 30.5% of zones are completely located within the boundaries of one census tract. Forty-eight percent of zones straddle two or three census tracts whereas, 21.5% spread over more than four tracts. Correspondingly, there is considerable variability in types of environment within census tracts.

Figure 5. Proportion of zones straddling different numbers of census tracts across the Island of Montreal



As illustrated in Figure 6, 11.2% of census tracts encompass only one type of environment and 34.3% encompass two types. About 28% of census tracts are characterised by three different types of environment, whereas 26.3% comprise 4 or more different types. Among census tracts encompassing two types of environment (n=175), about two-thirds (66.3%) comprise types that are statistically similar as indicated by distances between their centroids (two or less distance lag as indicated in the distance matrix in Figure 3; results not shown). For example, census tracts often comprise a combination of low-density suburban and suburban/urban axial zones

Figure 6. Number of types of environment encompassed within census tract boundaries



(26.3%), or a grouping of diverse and high accessibility central urban areas (35.4%). Globally, the number of types of environment encompass within census tracts is positively correlated in space (Moran I = 0.26; p<0.001), suggesting that more homogeneous or more mixed census tracts are often contiguous in space (Figure 6). More heterogeneous census tracts are located mainly on the periphery of central urban areas and in the eastern part of the Island of Montreal, and to a lesser extend in the west-end suburbs.

DISCUSSION

The objective of this study was to assess the soundness of census tracts as units of analysis, i.e. their degree of homogeneity in terms of the active living potential of residential environments associated with walking. In order to do so, homogeneous zones that optimised the spatial patterning of active living potential indicators hypothesised to be associated with greater involvement in walking, namely population density, land-mix use, and accessibility to services, were successfully designed. This was done through application of an easy-to-use method combining a classification method called *K*-means clustering with basic mapping applications of geographical information systems. The degree of soundness of census tracts as units of analysis was established through a series of analyses comparing them to the newly-designed zones.

First the distribution of the three active living indicators between and within census tracts was assessed. Although census tracts were homogeneous in terms of accessibility to services, they were less homogenous in population density; for this indicator within and between census tracts variations were about equal. Census tracts were clearly not homogeneous in terms of land use mix as the variability within tracts largely exceeded the variability between tracts. In contrast, census tracts were homogeneous along socioeconomic variables. These results suggest that the spatial patterning of the active living potential of environments do not neatly follow in the delineation of census tracts, which may be more suitable as units of analysis for operationalising socioeconomic contexts.

Then, findings revealed that the spatial distribution of active living and socioeconomic indicators followed different spatial distribution. At the zone-level, types of environment explained a small proportion of variation of socioeconomic variables. This indicates that processes underlying the distribution of active living and SEP indicators, although potentially linked [2, 6], operate at different scales and thus require different units of analysis.

In the final set of analyses, within tract variability in terms of what we labelled "types of environment" was examined. This allowed for the assessment of whether or not census tracts encompassed environments that were substantively different among them in terms of their active living potential. Census tracts comprising two different types of environments (34.3%) were not considered necessarily as problematic, given that some types of environment were more similar than others and were often contiguous in space. For example, diverse and high accessibility central urban zones were often contiguous in space and were statistically most similar (as indicated by statistical distances between clusters; Figure 3). However, census tracts comprising three or more types of environment raised concerns; such a situation was observed in more than half of census tracts. These tracts encompass environments that are simultaneously most conducive to walking and others that are least so. Averaging values of conduciveness to walking could potentially lead to significant errors when measuring active living potential at the census tract-level.

The approach for defining areas or units of analysis differs from those involving the definition of strictly "ecologically meaningful" or "natural" neighbourhoods, i.e. neighbourhoods imbued with meaning for residents [21] or as consisting of a group of homes sharing a commonly defined residential area often having name [20]. Defining such units of analysis is important when the notion of commonly shared territory is related to the contextual condition of interest, for example social capital or collective efficacy [7, 27]; this notion is not conjured up by active living potential. Designing zones based on the spatial distribution of active living indicators empirically linked to greater involvement in walking leads to the definition of areas that are more appropriate units of analysis and increases the internal validity of study design examining the environmental determinants of walking.

Future studies are needed to assess the impact of the choice of other environmental characteristics for designing zones relevant to other health indicators, and to other geographical areas. For example, areas relevant for studying the social and environmental determinants of overweight and obesity may be delimited according to the distribution of active living variables and food provision (accessibility of both healthy and non-healthy food). For studying mental health outcomes, social dimensions of area context, such as social support and opportunities for social participation, may be more relevant. It is to be expected that designing zones using other indicators of contextual conditions associated with other health outcomes will lead to different spatial configuration of area units of analysis.

Homogeneous zones are designed with the aim of optimising the study of a phenomenon or for the purpose of uncovering the aetiology underlying associations between area context and health. As such, the configuration of zones should not be viewed as other "spaces" of actions for public health and policy interventions. Rather, they may be useful for informing on viable interventions and policy strategies that may be health promoting.

Limitations

Results of this study should be considered in light of some limitations. First, there is a seven year time lag (2000 to 2006) between the dates of creation of the different datasets used to characterise dissemination areas in terms of active living potential and socioeconomic position. Although changes may have occurred during this period, the speed at which changes in the built environment occur is not well documented; however over a five-year course, changes in the built environment can be expected to be modest.

Other indicators of active living potential could be examined in designing homogeneous areas, such as street connectivity, safety, and accessibility to other services or resources such as parks. In this study, the measurement of land use mix was dependent on the size of disseminations areas which are defined in part by a population size threshold; they are likely to span a greater territory when located in suburban areas in comparison to urban areas, where population density is higher, and therefore encompass more land use. Other scales for measuring land use mix could be considered [47].

CONCLUSION

For studies concerned with the social and environmental determinants of health and more specifically of physical activity, results of this study have several implications. Delimiting areas is a key conceptual and methodological challenge in research on health and place. In this paper, we developed an easy-to-use method for establishing homogeneous units of analysis in terms of specific environmental characteristics hypothesised to be linked to a specific health indicator. The focus was on active living potential of areas and walking behaviours. Using these homogeneous zones as comparison, the objective was to assess the soundness of spatial units "of convenience", i.e. census tracts, to operationalise contexts for which they were not purposely developed. The methods developed in this study add to the growing literature on alternative ways to conceptualise and define the boundaries of area units for studying the determinants of health.

Findings showed that although census tracts may be homogeneous along independent indicators of active living potential, they were most often characterised by a combination of types of environment that were substantively different in terms of their active living potential. For this reason, census tracts should be used with caution as units of analysis when operationalising active living potential for studying determinants of walking. But census tracts or other administratively defined areas may be appropriate area units, i.e. may be homogeneous enough, when processes hypothesised to be operating on health are linked to the socioeconomic context of an area, for example affluence or poverty.

In this study, zones were delimited for methodological and aetiological purposes with the aim of minimising measurement errors of environmental characteristics and increasing internal validity of study design for measuring area effects on health. As can be expected, the zones are context-specific and cannot be exported to other geographic areas. Rather they are representations of the local realities of processes relating environmental characteristics to health. As suggested by others, the geographical aspects of the study design should be considered prior to conducting analyses [15]. Establishing the soundness of spatial units "of convenience" for representing the environmental and spatial processes under investigation should be part of the empirical approach for conceptualising, operationalising, and measuring area effects on health.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

MR conceptualised the study. She carried out spatial and statistical analyses, mapping of results, and drafted the manuscript. PA, LG and JMB participated in the conceptualisation the study, and in data analyses. All authors critically revised the paper, and read and approved the final manuscript.

Acknowledgements

MR is the recipient of a Canada Graduate Scholarships Doctoral Award from the Canadian Institutes of Health Research.

References

1. Curtis S, Jones IR: **Is there a place for geography in the analysis of health inequality?** *Sociology of Health & Illness* 1998, **20**(5):645-672.
2. Macintyre S, Ellaway A, Cummins S: **Place effects on health: how can we conceptualise, operationalise and measure them?** *Social Science & Medicine* 2002, **55**(1):125-139.
3. Ellen IG, Mijanovich T, Dillman K-N: **Neighborhood Effects on Health: Exploring the Links and Assessing the Evidence**. *Journal of Urban Affairs* 2001, **23**(3&4):391-408.
4. Pickett KE, Pearl M: **Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review**. *Journal of Epidemiology and Community Health* 2001, **55**(2):111-122.
5. Riva M, Gauvin L, Barnett TA: **Toward the next generation of research into small area effects on health: A synthesis of multilevel investigations published since July 1998**. *Journal of Epidemiology and Community Health* 2007, **61**(10):853-861.
6. Daniel M, Moore S, Kestens Y: **Framing the biosocial pathways underlying associations between place and cardiometabolic disease**. *Health and Place* 2008, **14**(2):117-132.
7. Diez Roux AV: **Investigating neighborhood and area effects on health**. *American Journal of Public Health* 2001, **91**(11):1783-1789.
8. Blakely TA, Lochner K, Kawachi I: **Metropolitan area income inequality and self-rated health - a multi-level study**. *Social Science & Medicine* 2002, **54**(1):65-77.
9. Cockings S, Martin D: **Zone design for environment and health studies using pre-aggregated data**. *Social Science & Medicine* 2005, **60**(12):2729-2742.
10. Franzini L, Spears W: **Contributions of social context to inequalities in years of life lost to heart disease in Texas, USA**. *Social Science & Medicine* 2003, **57**(10):1847-1861.

11. Hou F, Myles J: **Neighbourhood inequality, neighbourhood affluence and population health**. *Social Science & Medicine* 2005, **60**(7):1557-1569.

12. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R: **Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project**. *American Journal of Epidemiology* 2002, **156**(5):471-482.

13. Reijneveld SA, Verheij RA, de Bakker DH: **The impact of area deprivation on differences in health: does the choice of the geographical classification matter?** *Journal of Epidemiology and Community Health* 2000, **54**(4):306-313.

14. Oliver L, Hayes MV: **Does choice of spatial unit matter for estimating small area disparities in health and place effects in the Vancouver census metropolitain area?** *Canadian Journal of Public Health* 2007, **98**:S27-S34.

15. Flowerdew R, Manley DJ, Sabel CE: **Neighbourhood effects on health: Does it matter where you draw the boundaries?** *Social Science & Medicine* 2008, **66**:1241-1255.

16. Chaix B, Chauvin P: **Tobacco and alcohol consumption, sedentary lifestyle and overweightness in France: A multilevel analysis of individual and area-level determinants**. *European Journal of Epidemiology* 2003, **18**(6):531-538.

17. Chaix B, Guilbert P, Chauvin P: **A multilevel analysis of tobacco use and tobacco consumption levels in France - Are there any combination risk groups?** *European Journal of Public Health* 2004, **14**(2):186-190.

18. Chuang YC, Cubbin C, Ahn D, Winkleby MA: **Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking**. *Journal of Epidemiology and Community Health* 2005, **59**(7):568-573.

19. van Lenthe FJ, Brug J, Mackenbach JP: **Neighbourhood inequalities in physical inactivity: the role of neighbourhood attractiveness, proximity to**

local facilities and safety in the Netherlands. *Social Science & Medicine* 2005, **60**(4):763-775.

20. Ross NA, Tremblay S, Graham K: **Neighbourhood influences on health in Montreal, Canada.** *Social Science & Medicine* 2004, **59**(7):1485-1494.
21. Coulton CJ, Korbin J, Chan T, Su M: **Mapping residents' perceptions of neighborhood boundaries: A methodological note.** *American Journal of Community Psychology* 2001, **29**(2):371-383.
22. Diez-Roux AV: **Multilevel analysis in public health research.** *Annual Review of Public Health* 2000, **21**:171-192.
23. Gauvin L, Robitaille E, Riva M, McLaren L, Dassa C, Potvin L: **Conceptualizing and operationalizing neighbourhoods: the conundrum of identifying territorial units.** *Canadian Journal of Public Health* 2007, **98**(suppl 1):S18-S26.
24. O'Campo P: **Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health.** *American Journal of Epidemiology* 2003, **157**(1):9-13.
25. Subramanian S, Jones K, Duncan C: **Multilevel methods for public health research.** In: *Neighborhoods and health.* Edited by Kawachi I BL. New York: Oxford University Press; 2003: 65-111.
26. Subramanian SV: **The relevance of multilevel statistical methods for identifying causal neighborhood effects - Commentary.** *Social Science & Medicine* 2004, **58**(10):1961-1967.
27. Osypuk TL, Galea S: **What level macro? Choosing appropriate level to assess how place influences population health.** In: *Macrosocial determinants of population health.* Edited by Galea S. New York: Springer; 2007: 399-438.
28. Cummins S, Macintyre S, Davidson S, Ellaway A: **Measuring neighbourhood social and material context: generation and interpretation of ecological data from routine and non-routine sources.** *Health & Place* 2005, **11**(3):249-260.

29. Gauvin L, Richard L, Craig CL, Spivock M, Riva M, Forster M, Laforest S, Laberge S, Fournel MC, Gagnon H *et al*: **From walkability to active living potential - An "ecometric" validation study**. *American Journal of Preventive Medicine* 2005, **28**(2):126-133.

30. Chaix B, Rosvall M, Lynch J, Merlo J: **Disentangling contextual effects on causespecific mortality in a longitudinal 23-year follow-up study: impact of population density or socioeconomic environment?** *International Journal of Epidemiology* 2006, **35**(3):633-643.

31. Galea S, Ahern J: **Invited commentary: Considerations about specificity of associations, causal pathways, and heterogeneity in multilevel thinking.** *American Journal of Epidemiology* 2006, **163**(12):1079-1082.

32. Openshaw S: **The modifiable areal unit problem**. *Concepts and Techniques in Modern Geography* 1984, **38**:1-41.

33. Openshaw S, Taylor PJ: **A Million or So Correlated Coefficients: Three experiments on the Modifiable Areal Unit Problem**. In: *Statistical applications in the spatial sciences.* Edited by Wrigley N, Bennet R. London: Pion; 1979: 127-144.

34. Cliff A, Ord J: **Spatial autocorrelation**. In: *Monograph No 5 in spatial and environmental systems analysis.* Edited by Chorley R, Harvey D. London: Pion Ltd; 1973: 178.

35. Cummins S, Curtis S, Diez-Roux AV, Macintyre S: **Understanding and representing 'place' in health research: A relational approach**. *Social Science & Medicine* 2007, **65**(9):1825-1838.

36. Rothman K, Greenland S: **Modern epidemiology**, 2nd edn. Philadelphia: Lippincott Williams & Wilkins; 1998.

37. Haynes R, Daras K, Reading R, Jones A: **Modifiable neighbourhood units, zone design and residents ' perceptions**. *Health & Place* 2007, **13**(4):812-825.

38. Haynes R, Gale S: **Mortality, long-term illness and deprivation in rural and metropolitan wards of England and Wales**. *Health & Place* 1999, **5**(4):301-312.

39. Blakeley T, Subramanian SV: **Multilevel studies**. In: *Methods in social epidemiology*. Edited by Oakes JM, Kaufman JS. San Francisco, CA: Jossey-Bass; 2006: 316-340.

40. Statistics Canada Census Operations Division: **2001 Census Dictionary**. In. Ottawa, Canada: Minister of Industry; 2003.

41. Gauvin L, Riva M, Barnett TA, Richard L, Craig CL, Spivock M, Laforest S, Laberge S, Fournel MC, Gagnon H *et al*: **Association between neighborhood active living potential and walking**. *American Journal of Epidemiology* 2008, **167**:944-953.

42. Cervero R: **Mixed land-uses and commuting: Evidence from the American housing survey**. *Transportation Research Part a-Policy and Practice* 1996, **30**(5):361-377.

43. De Bourdeaudhuij I, Sallis JF, Saelens BE: **Environmental correlates of physical activity in a sample of Belgian adults**. *American Journal of Health Promotion* 2003, **18**(1):83-92.

44. Ewing R, Schmid T, Killingsworth R, Zlot A, Raudenbush S: **Relationship between urban sprawl and physical activity, obesity, and morbidity**. *American Journal of Health Promotion* 2003, **18**(1):47-57.

45. Fisher K, Li F, Michael Y: **Neighborhood-level influences on physical activity among older adults: a multi-level analysis**. *Journal of Aging and Physical Activity* 2004, **12**(1):45-63.

46. Frank LD, Pivo G: **Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking.** *Transportation Research Record* 1995, **1466**:44-52.

47. Frank LD, Schmid TL, Sallis JF, Chapman J, Saelens BE: **Linking objectively measured physical activity with objectively measured urban form - Findings from SMARTRAQ**. *American Journal of Preventive Medicine* 2005, **28**(2):117-125.

48. Giles-Corti B, Broomhall MH, Knuiman M, al. e: **Increasing walking: how important is distance to, attractiveness, and size of public open space?** *American Journal of Preventive Medicine* 2005, **28**(Suppl 2):169-176.

49. Giles-Corti B, Donovan RJ: **Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment**. *Preventive Medicine* 2002, **35**(6):601-611.

50. Giles-Corti B, Donovan RJ: **Relative influences of individual, social environmental, and physical environmental correlates of walking.** *American Journal of Public Health* 2003, **93**(9):1583-1589.

51. Lee C, Vernez Moudon A: **Correlates of Walking for Transportation or Recreation Purposes**. *Journal of Physical Activity and health* 2006, **3** (Suppl 1):S77.

52. Pikora TJ, Giles-Corti B, Knuiman MW, Bull FC, Jamrozik K, Donovan RJ: **Neighborhood environmental factors correlated with walking near home: Using SPACES**. *Medicine and Science in Sports and Exercise* 2006, **38**(4):708-714.

53. Saelens BE, Sallis JF, Frank LD: **Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures**. *Annals of Behavioral Medicine* 2003, **25**(2):80-91.

54. Eyler AA, Brownson RC, Bacak SJ, Housemann RA: **The epidemiology of walking for physical activity in the United States**. *Medicine and Science in Sports and Exercise* 2003, **35**(9):1529-1536.

55. Rafferty AP, Reeves MJ, McGee HB, Pivarnik JM: **Physical activity patterns among walkers and compliance with public health recommendations**. *Medicine and Science in Sports and Exercise* 2002, **34**(8):1255-1261.

56. United States Department of Health and Human Services: **Physical Activity and health: A report from the Surgeon General.** In. Atlanta, GA: U.S: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 1996.

57. Browning CR, Cagney KA: **Moving beyond poverty: Neighborhood structure, social processes, and health**. *Journal of Health and Social Behavior* 2003, **44**(4):552-571.

58. Lochner KA, Kawachi I, Brennan RT, Buka SL: **Social capital and neighborhood mortality rates in Chicago**. *Social Science & Medicine* 2003, **56**(8):1797-1805.

59. Reading R, Langford IH, Haynes R, Lovett A: **Accidents to preschool children: comparing family and neighbourhood risk factors**. *Social Science & Medicine* 1999, **48**(3):321-330.

60. Law M, Wilson K, Eyles J, Elliott S, Jerrett M, Moffat T, Luginaah I: **Meeting health need, accessing health care: the role of neighbourhood**. *Health & Place* 2005, **11**(4):367-377.

61. Lebel A, Pampalon R, Villeneuve PY: **A multi-perspective approach for defining neighbourhood units in the context of a study on health inequalities in the Quebec City region**. *International Journal of Health Geographics* 2007, **6:27**.

62. Martin D, Nolan A, Tranmer M: **The application of zone-design methodology in the 2001 UK Census**. *Environment and Planning A* 2001, **33**(11):1949-1962.

63. Nakaya T: **An information statistical approach to the modifiable areal unit problem in incidence rate maps**. *Environment and Planning A* 2000, **32**(1):91-109.

64. Openshaw S: **Developing GIS-relevant zone-based spatial analysis methods**. In: *Spatial analysis: Modelling in a GIS environment.* Edited by Longley P, Batty M. Cambridge: Pearson Professional Ltd; 1996: 55-73.

65. Openshaw S, Rao L: **Algorithms for reengineering 1991 census geography**. *Environment and Planning A* 1995, **27**(3):425-446.

66. **Ville de Montréal** [www.ville.montreal.qc.ca]

67. ESRI: **ArcGIS 9.2**. In. Redlands, CA; 2006.

68. Statistics Canada: **2001 Census preview of products and Services.** In. Ottawa, Canada: Statistics Canada.

69. Theil H: **Statistical decomposition analysis**. Amsterdam: North-Holland; 1972.

70. Theil H, Finezza A: **A note on the measurement of racial integration of schools by means of informational concepts**. *Journal of Mathematical Sociology* 1971, **1**:187-194.

71. Communauté urbaine de Montréal: **Carte d'occupation du sol (édition 2000); document d'accompagnement**. Montréal: Communauté urbaine de Montréal, division de l'aménagement, service de la mise en valeur du territoire; 2001.

72. Cloutier MS, Apparicio P, Thouez JP: **GIS-based spatial analysis of child pedestrian accidents near primary schools in Montréal, Canada**. *Applied GIS* 2007, **3**(4).

73. Hewko J, Smoyer-Tomic KE, Hodgson MJ: **Measuring neighbourhood spatial accessibility to urban amenities: Does aggregation error matter?** *Environment and Planning A* 2002, **34**(7):1185 - 1206.

74. Witten K, Exeter D, Field A: **The quality of urban environments: Mapping variation in access to community resources**. *Urban Studies* 2003, **40**(1):161-177.

75. Apparicio P, Séguin AM: **Measuring the accessibility of services and facilities for residents of public housing in Montréal**. *Urban Studies* 2006, **43**(1):187-211.

76. Apparicio P, Abdelmajid M, Riva M, Shearmur R: **Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues**. *International Journal of Health Geographics* 2008 **7**:7.

77. DMTI Spatial: **CanMap® Streetfiles**. Markham: DMTI Spatial inc.; 2005.

78. Tabachnick B, Fidel L: **Using multivariate statistics** 4th edn. Needham Heights, MA: Allyn & Bacon; 2001.

79. SAS Institute Inc: **SAS version 9.1**. In. Cary, NC, USA.

80. Harris R, Sleight P, Webber R: **Geodemographics, GIS and neighbourhood targetting**. Chichester, West Sussex, England: John Wiley & Sons, Ltd; 2005.

81. Duda R, Hart P, Stork D: **Pattern classification**, 2 edn. New York: John Wiley & Sons, Ltd; 2001.

82. Everitt B, Landau S, Leese M: **Cluster analysis.**, 4th edn. London: Arnold; 2001.

83. Lebart L, Morineau A, Piron M: **Statistique exploratoire multidimensionnelle**, 2nd edn. Paris: Dunod; 1997.

84. Raudenbush S, Bryk A, Congdon R: **HLM: Hierarchical Linear and Nonlinear Modeling (Version 6.04)** In. Chicago: Scientific Software International; 2005.

85. Lee J, Wong DWS: **Statistical analysis with ArcView GIS**. New York: John Wiley & Sons, Inc; 2001.

86. Longley P, Batty M: **Spatial analysis: Modelling in a GIS environment**. Cambridge: earson Professional Ltd; 1996.

87. Milligan GW, Cooper MC: **An examination of procedures for determining the number of clusters in a data set.** *Psychometrika* 1985, **50**(2):159-179.

88. Calinski T, Harabasz J: **A dendrite method for cluster analysis**. *Communications in Statistics* 1974, **3**(1):1-27.

89. SAS Institute Inc: **SAS Technical Report A-108, Cubic clustering criterion.** . In. Cary (NC), USA: SAS Institute Inc: 54.

ABSTRACT

The aim of this study was to examine the contribution of designing homogenous spatial units along the exposure of interest for studying area effects on health, using the examination of the association between active living potential and walking as an example. Zone design ensued from the statistical classification of dissemination areas into seven optimal categories of exposure based on similarity along three indicators of active living potential: population density, land use mix, and geographical accessibility to proximity services. Classifications were then imported into a geographical information system resulting in the creation of zones which were homogenous in terms of categories of exposure. Active living potential and socioeconomic indicators were also computed at the census tract level. Individual data on walking behaviours among a sample of 2716 adults aged 45 years and older, were nested within 112 census tracts and 270 zones and analysed using two-level and cross-classified multilevel models. Findings show that accessibility to services appears to be an important predictor of walking. Although the magnitude of this effect is larger at the zone-level, estimates were similar to those observed across census tracts. Nonetheless using area units homogenous in exposures increases the internal validity of the study design and therefore improves the soundness of statistical inference. Results of cross-classified models wherein active living and socioeconomic exposures were measured in different spatial units, i.e. respectively zones and census tracts, draw attention to the diverse overlapping contexts influencing walking. Designing spatial units maximising homogeneity of exposures is advantageous in aetiological research as it strengthens internal validity without becoming unwieldy. Replication studies in relation to other health indicators are required to further assess the value of the novel method for designing zones relevant.

INTRODUCTION

In area effects on health research, although the validity and reliability of ecologic exposure measures are being perfected (Cummins, Macintyre, Davidson, & Ellaway, 2005; Frank, Schmid, Sallis, Chapman, & Saelens, 2005; Oakes & Rossi, 2003; Pearce, Witten, & Bartie, 2006; Raudenbush & Sampson, 1999), defining areas remains a challenge that gives rise to conceptual and empirical issues potentially challenging the construct and internal validity of studies (Osypuk & Galea, 2007).

Several authors have developed novel approaches for conceptualising and operationalising area units (Cockings & Martin, 2005; Frohlich, Potvin, Chabot, & Corin, 2002; Haynes, Daras, Reading, & Jones, 2007; Popay, Thomas, Williams, Bennett, Gattrell, & Bostock, 2003). These efforts are far too few in comparison to studies operationalising areas using administrative spatial units such as census tracts, wards, boroughs, or postcode sectors (Pickett & Pearl, 2001; Riva, Gauvin, & Barnett, 2007). These units are useful because they can easily be linked to data from censuses and other surveys that can be used to derive ecologic exposure measures. Additionally, as administrative units are often designed to be homogeneous along socioeconomic dimensions, they may be appropriate for operationalising socioeconomic contexts (Ross, Tremblay, & Graham, 2004). However, other contextual dimensions may not be optimally defined and measured within these readily available spatial units.

To date, few studies have explored the soundness of administrative spatial units for research on area effects on health (Cockings & Martin, 2005; Haynes et al., 2007). For understanding the significance of areas for health, it is important that areas be defined in coherence with the specific objectives of the study in terms of health indicators, exposures, or their association, is important (Chaix, Leyland, Sabel, Chauvin, Rastam, Kristersson et al., 2006; Cummins et al., 2005; Galea & Ahern, 2006; Gauvin, Robitaille, Riva, McLaren, Dassa, & Potvin, 2007) as these would allow for the unbiased measurement of exposure and their influence on health.

Toward this end, establishing the validity of spatial units, i.e. the homogeneity od spatial units in the exposures of interest, is a major methodological consideration.

In a previous study set on the Island of Montreal, Canada (Riva, Apparicio, Gauvin, & Brodeur, In Press), we assessed the soundness of census tracts for studying associations between walking behaviours and the active living potential of the built environment of residential areas, that is the conditions of areas that encourage the likelihood of integrating physical activity into daily routines (Gauvin, Richard, Craig, Spivock, Riva, Forster et al., 2005). In this latter investigation, active living potential was operationalised with three indicators, namely population density, land use mix, and geographic accessibility to proximity services. In order to do so, we designed 898 homogeneous zones characterised by one of seven categories of exposure resulting from the interplays between the three indicators in creating distinct active living potentials. Homogeneity of census tracts (n=506 with valid population and socioeconomic data) was assessed in comparison to the newly-designed zones.

We observed that although census tracts were homogenous with respect to accessibility to services, they were heterogeneous in terms of population density and land use mix. Furthermore, census tracts were most often characterised by a combination of categories of exposure that were substantively different from one another. That is, more than half of census tracts studied (55.5%) encompassed environments with different levels of active living potential. These results indicate that the soundness of census tracts for studying the environmental determinants of walking may be limited. Furthermore, measuring active living exposure variables at the census tract level may lead to exposure misclassification and misestimation of associations therefore threatening the internal validity of the study design by limiting the soundness of inference about area effects on walking.

In contrast, census tracts were more homogeneous in relation to socioeconomic characteristics. However, the distribution of socioeconomic and active living exposure variables followed different spatial distributions, suggesting that different spatial units may be relevant for different exposures. In research on area effects, the socioeconomic and built environments are often measured within the same spatial units, e.g. census tracts, wherein greater homogeneity along socioeconomic dimensions is more likely. Findings of this study suggest that the strategy that consists of designing area units maximizing homogeneity of exposure variables may be appropriate for addressing the methodological challenge of operationalising areas.

Study objectives

In order to examine the contribution of designing spatial units homogenous on the exposure of interest, the objectives of this study are twofold: 1) in separate multilevel models, to examine and compare results of associations between area-level active living potential and walking behaviours when areas are operationalised by the newly-defined zones and by census tracts; and 2) in cross-classified multilevel models, to assess the relative influence of active living potential and socioeconomic exposure variables on walking behaviours using different area units that are homogeneous along these exposures, respectively zones and census tracts.

METHODS

Sampling

The study area is the Island of Montreal, Canada, an urban centre with 1 812 723 inhabitants. Individual data on walking behaviour and socio-demographic characteristics are from a larger project aimed at understanding the environmental determinants of walking in Montreal, Canada (Gauvin et al., 2005; Gauvin, Riva, Barnett, Richard, Craig, Spivock et al., 2008). Stratified sampling of participants resulted from random samples of census tracts and of individuals within these tracts.

At the time of data collection in 2005, the Island of Montreal was divided into 27 boroughs[1] and 521 census tracts. Because census tracts were unevenly nested within boroughs (min=4, max=48 census tracts per borough), they were categorized based on tertiles of average household income. A random sample of 20% of tracts within each borough strata was selected, with the constraint that at least two census tracts per borough were selected. The final sample comprised 112 census tracts.

Participants living in the 112 census tracts were randomly recruited by a recognised polling firm between February and May 2005. Respondents were invited to participate in a twenty-minute telephone interview. There were four eligibility criteria: (1) being aged 45 years or older; (2) living in one of the 112 census tracts as confirmed by the verification of the 6 digit postal code; (3) having lived at the current address for at least one year; and (4) being able to respond in either French or English. Only one respondent per household participated. A sample of 2923 (response rate=29.8%) participated in the survey. After listwise deletion due to incomplete or missing responses, data from 2716 (92.9%) participants were analysed.

Individual measures

Walking. In research on area effects on health, understanding the environmental determinants of walking is receiving increasing attention. Practiced regularly, i.e. 30 to 60 minutes per day on most days of the week, walking translates into significant health benefits (Lee & Paffenbarger, 2000; Pate, Pratt, Blair, Haskell, Macera, Bouchard et al., 1995; USDHHS, 1996). When sedentary people become moderately active, health benefits increase at a rapid rate and then taper off when people advance from being moderately active to very active (USDHHS, 1996). Walking requires low level of exertion, imposes few barriers (time, monetary, or accessing facilities); it is the most common type of physical activity across all age, income, and ethnic group (King, Brach, Belle, Killingsworth, Fenton, & Kriska, 2003; Siegel, Brackbill, &

[1] 1. As of January 2006, following merging/de-merging of boroughs into one mega-city of Montreal, there are 15 municipalities on the Island of Montreal, in addition to the municipality of Montreal which includes 19 boroughs; www.ville.montreal.qc.ca

Heath, 1995). For these reasons, promoting regular walking is one strategy to address the public health burden of physical inactivity (Transportation Research Board and Institute of Medicine of the National Academies (TRB & IOM), 2005; United States Department of Health and Human Services (USDHHS), 1996; World Cancer Research Fund / American Institute for Cancer Research (WCRF / AICR), 2007).

Walking was assessed using questions from the International Physical Activity Questionnaire, a 7 item questionnaire with good test-retest reliability (r=0.80) and moderate convergent validity (r=0.30) with accelerometers (Craig, Marshall, Sjostrom, Bauman, Booth, Ainsworth et al., 2003). Respondents indicated on how many days they walked for any motive (e.g., walking to get around, to maintain health) for at least 10 minutes at a time in the previous 7 days. Those who walked on at least one day were asked to estimate the amount of time per day spent walking. Additional questions addressed walking for recreational purposes. Participants who reported walking during the previous week were asked to indicate if any walking episodes were performed specifically to maintain health/fitness. Those responding affirmatively were asked to estimate the number of days and typical duration. Total weekly minutes of walking for any motive and for recreational purposes were computed by multiplying number of minutes walked per day by the number of days of walking. We created a residual walking variable, walking for utilitarian purposes, defined by subtracting the total weekly minutes of recreational walking from that of walking for any motive. Although it would have been ideal to address a set of questions specifically on utilitarian walking in the telephone interview, pilot testing showed that such questioning over the telephone was unwieldy and resulted in participant confusion.

From these variables, three dependent variables were computed: the number of 10-minute episodes of walking for any motive, for recreational reasons, and for utilitarian purposes as obtained by dividing by ten the total weekly minutes of walking for each motive. These measures were selected because of studies showing health benefits when people accumulate bouts of 10 minutes of physical activity throughout the day (USDHHS, 1996).

Individual socio-demographic characteristics. Potential individual socio-demographic confounders accounted for were sex, age, and educational attainment from which dummy variables were created. Age was computed from participants' reported birth year and re-categorized as between 45 and 54 years, 55 to 64, or 65 years and older. Participants reported their highest academic degree: less than high school, high school diploma, trade diploma or college diploma, and university degree completed.

Active living potential exposure variables and zone design

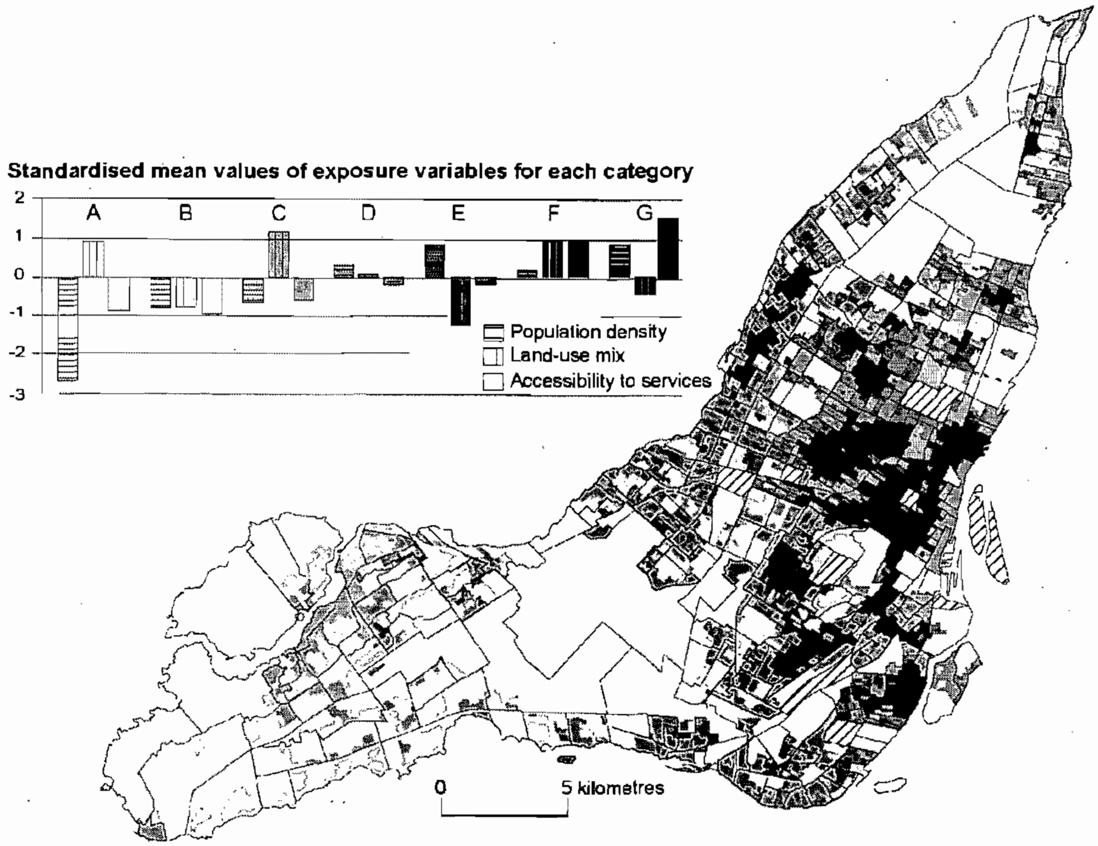
Zones homogeneous along active living potential were designed based on the spatial distribution of three exposure variables linked to the concept of density of destinations, i.e. the physical and social characteristics of residential areas related to land use patterns (Gauvin et al., 2005), namely population density, land use mix, and accessibility to services (Riva et al., In Press). These variables were selected because they were observed to be significant predictors of increased levels of walking in several studies (Cervero, 1996; De Bourdeaudhuij, Sallis, & Saelens, 2003; Fisher, Li, & Michael, 2004; Frank & Pivo, 1995; Frank et al., 2005; Gauvin et al., 2008; Giles-Corti, Broomhall, Knuiman, & al., 2005; Giles-Corti & Donovan, 2002a, b; Humpel, Owen, Leslie, Marshall, & Bauman, 2004; C. Lee & Vernez Moudon, 2006; Li, Fisher, Brownson, & Bosworth, 2005; Owen, Humpel, Leslie, Bauman, & Sallis, 2004; Pikora, Giles-Corti, Knuiman, Bull, Jamrozik, & Donovan, 2006; Saelens, Sallis, & Frank, 2003).

In a first step, values for population density (total number of residents / km^2), land use mix, and geographical accessibility to selected proximity services were computed at the dissemination area level (DAs; n=3206) using ArcGIS 9.2 (ESRI, 2006). Dissemination areas were used as building blocks because they are the smallest standard geographic area for which Canadian census data are available. Land use mix was computed using an entropy index (Theil, 1972; Theil & Finezza, 1971) measuring diversity of a component, e.g. land use, within a spatial unit. Geographical accessibility (hereafter referred to as accessibility) to proximity services, i.e. to supermarkets, pharmacies, banks, and libraries, was defined by the number of these services within a one kilometre network buffer (radius delimited by the road network) computed from the centroid of census blocks comprised within any one DA (this was done to minimise aggregation errors); values were averaged to the DA-level. Using K-means statistical clustering, the 3206 DAs were optimally classified into seven distinct homogeneous "categories of exposure". The approach was successful in maximising both differences across the seven categories of exposure and similarity of DAs within one category, i.e. within-cluster homogeneity. The categories of exposures were then mapped into ArcGIS 9.2, leading to the delineation of 898 homogeneous zones characterised by one of the seven categories of exposure.

The seven categories of exposure encompassed more suburban to more central urban types of built environments defined by different values for population density, land use mix, and accessibility to services; the categories of exposures are described in Figure 1. Zones characterised by more suburban contexts were on average larger and had relatively smaller population counts than central urban zones. Low-density and mid-density suburban areas are characterised by lower values for population density and accessibility to services, whereas diverse central urban areas and central urban areas with very good accessibility to services are more densely populated and have greater access to services than any other categories of exposures.

Analyses were conducted on a sample of 270 zones (30.1%) wherein the 2716 individuals were geocoded (using their six-digit postal codes and match up files). The seven categories of exposure were the main exposure variables and were modelled as dummy variables; low-density suburban zone was the reference category. Zone-level socioeconomic context was also computed whereby dissemination area data on the proportion of people aged 20 years and older with university degree (2001 Canadian census) were aggregated at the zone-level (Statistics Canada). Zones were

Figure 1. Zones and categories of exposure



Description of categories of exposure			Zones N	Zones %	Area km2	Area %
A	Low-density suburban	Suburban areas characterised by lowest population density, land use diversity, and low accessibility to services	31	3.5	222.2	45.5
B	Middle-density suburban	Suburban areas characterised by low population density, land use homogeneity and weak accessibility to services	104	116	75.2	15.4
C	Suburban/ urban axial	Suburban and urban areas characterised by low population density, land use diversity and weak accessibility to services. Located close to main street axis	146	16.3	83.7	17.1
D	Mix urban/ suburban	Suburban and urban areas with average values of population density, land use mix, and accessibility to services. Located mainly in the central and eastern part of the island	243	27.1	41.0	8.4
E	Urban residential	Urban areas with highest population density, land use homogeneity, and weak accessibility to services. Located mainly in peripheral residential areas to central urban areas	212	23.6	17.1	3.5
F	Diverse central urban	Urban areas characterised average population density, land use diversity, and good accessibility to services	85	9.5	32.6	6.7
G	Central urban with high accessibility	Urban areas with high population density, land use homogeneity, and highest accessibility to services. Located mainly in central areas of the Island of Montréal	77	8.6	16.5	3.4
			898	**100.0**	**488.3**	**100.0**

categorised in tertiles from which dummy variables were created; lower and higher tertiles of proportion of university graduates were contrasted to the middle tertile.

Census tracts

Census tracts are small, relatively stable geographic areas with population ranging between 2500 and 8000 persons and located within large cities having an urban core population of 50 000 or more. Census tracts are designed to be homogeneous in terms of socioeconomic characteristics, e.g. similar economic status and social living conditions (Statistics Canada Census Operations Division, 2003).

Measures of population density, land use mix, and accessibility to services were computed at the census tract level using the same approach as described above. Census tracts were categorised in tertiles from which dummy variables were created; the lower tertile was the reference category.

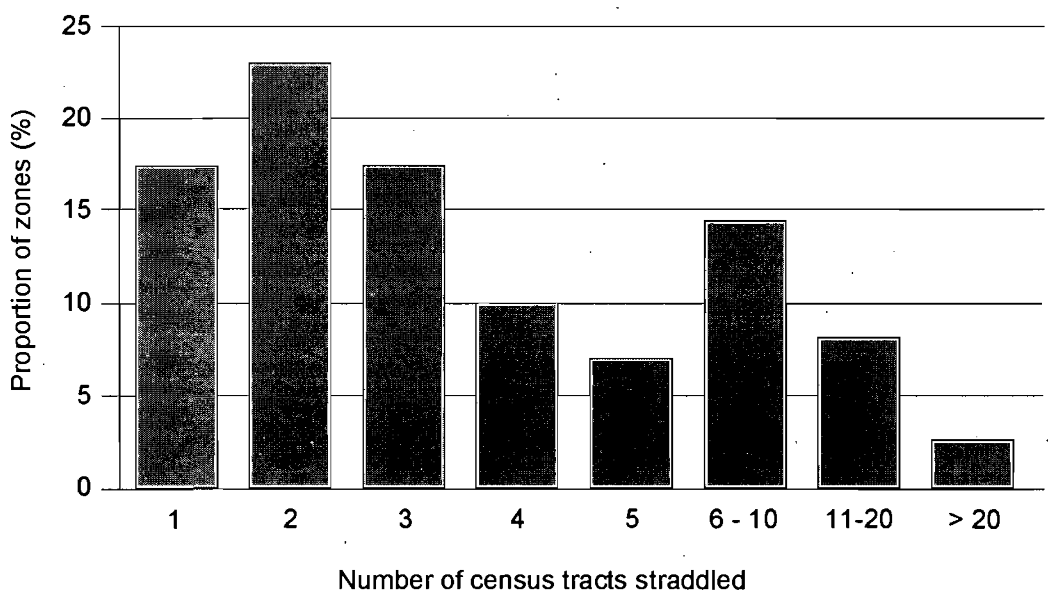
Census tract data pertaining to the proportion of persons with a university education were obtained from the 2001 Canadian census (Statistics Canada). The 112 census tracts were categorised in tertiles from which dummy variables were created; lower and higher tertiles of proportion of university graduates were contrasted to the middle tertile.

Statistical analyses

To examine and compare the predictive value of active living potential measured at the zone and census tract levels, three sets of multilevel analyses were performed. First to reflect conventional analyses procedures, we examined associations between active living and socioeconomic exposure variables and walking in a two-level model wherein individuals were nested within census tracts. Second, the same associations were examined with individuals nested within zones homogeneous in terms of the exposure of interest. Finally, to disentangle the relative influence of active living potential and socioeconomic status, data were analysed in cross-classified multilevel models. In these models, individuals were nested within two sets of overlapping area units operationalising different contexts: zones that were homogeneous in terms of active living potential and census tracts that are homogeneous along socioeconomic dimensions related to education.

Cross-classified analyses are more coherent with the structure of our spatial datasets. In our sample of 270 zones, 82.6% (n=223/270) straddled more that one census tracts (Figure 2). The spatial datasets were not hierarchically structured – rather they were crossed. As a result individuals living in the same specific zone could be exposed to different socioeconomic environments, or individuals living within the same specific socioeconomic environment may be exposed to a different active living potential. Since walking behaviours could plausibly be influenced by both dimensions, assessing their independent effects is pertinent.

Figure 2. Proportion of zones straddling different number of census tracts



For the three series of multilevel analyses, the modelling procedures followed a step-up approach. Null (unconditional) models were first performed to partition variance in the number of 10-minute walking episodes between individuals, census tracts, and zones. Associations between active living potential, socioeconomic status, and walking were estimated from three successive models: Model 1: Model with main exposure variables, i.e. the three active living exposure variables for census tracts, and the seven categories of exposure for zones; Model 2: Model with main exposure variables and area-level socioeconomic variable; Model 3: Model with main exposure variables and area-level socioeconomic variable, adjusted for the characteristics of individuals, to examine if associations between area-level factors and walking behaviour were attenuated by controlling for individual characteristics.

Associations between the number of events, i.e. 10-minute episodes of walking, and active living potential and socioeconomic status are reported by event rate ratios (ERR) and 95% confidence intervals (95% CI). Data were analysed using the Poisson sampling model and the log link function for count data with constant exposure (exposure corresponds to zones; for every resident of a zone, the exposure is the same) and adjusted to account for overdispersion (under the Poisson model, the variance and the mean are expected to be equal; in our data, variance was greater than the mean). Analyses were conducted in HLM software version 6.04 (Raudenbush, Bryk, & Congdon, 2005).

RESULTS

Description of individual, zone, and census tract samples

The individual sample included 1678 women (61.8%) and 1038 men (38.2%). Individuals aged between 45 and 54 years accounted for 28.7% of the sample, those 55 to 64 years for 34.7% whereas individuals aged 65 years and older represented 36.6%. About one third of the sample had completed university (31.4%) whereas 23.2% did not complete high school.

Average sample sizes within zones was 10 individuals (median 6; standard deviation: 12.4); 43.0% of zones comprised less than 5 individuals. In zones including five or more individuals, the proportion of individuals with annual household income below $20 000, with less than high school education, and with university education were correlated at 0.57 ($p<0.01$), 0.61 ($p<0.01$), and 0.72 ($p<0.01$) respectively with corresponding indicators obtained from Statistics Canada generated for each zone. This suggests that the within-zone (individual) sample is similar to the population living in these zones.

Zones were on average larger than census tracts (1.28 km^2 vs. 0.82 km^2), but this difference was not statistically significant ($t=0.78$; $p=0.438$). Zones and census tracts were also similar in terms of average population counts (3903 individuals in zones vs. 3715 in census tracts; $t=0.304$; $p=0.761$). Yet in comparison to census tracts, there was greater variability in zones in area size ($F=3.64$; $p=0.057$) and in population counts ($F=29.61$; $p<0.001$).

In terms of variation in exposure variables across and within the 112 census tracts, a greater proportion of variation in accessibility to services was attributable to differences between census tracts (83.7%), although for population density (57.1%) and land use mix (79.3%) there was greater within-census tract variability, suggesting that overall, tracts are heterogeneous along active living exposure variables. However, census tracts were homogeneous in terms of the proportion of university graduates, where a greater portion of the variation was attributable to difference between census tracts (76.4%). In terms of categories of exposures, the 112 census tracts were heterogeneous: about 45.5% of tracts encompassed one or two categories of exposures, 31.3% encompassed three, and 23.2% were characterised by four or five different categories of exposures (results not shown, , see Riva et al., In review).

Results of unconditional two-level and cross-classified multilevel models

Results of unconditional two-level and cross-classified models are shown in Table 1. In two-level models, walking for any motive significantly varied across census tracts ($p<0.05$) whereas for utilitarian walking there was greater variation across zones than across census tracts ($p<0.01$), although both variations were statistically significant. Whereas across all spatial units the average number of utilitarian walking episodes was 15 per week, these episodes varied between 8 and 29 across census tracts and between 7 and 35 across zones (as calculated from the plausible value range) (Raudenbush & Bryk, 2002).

Cross-classified unconditional models partition total variation in walking episodes between individuals and between level-2 units, which is the sum of variation between zones and between census tracts. As shown in Table 1, partitioning of variation in the number of 10-minute episodes of walking for any motive and for utilitarian reasons was significant across census tracts only. However, the extent of this variation warrants further attention. As observed in the two-level zone model, a greater proportion of level-2 variation in utilitarian walking was attributable to differences between zones (63.9%) rather than to census tracts (36.1%). The non significance of this variation in the cross-classified model likely relates to the small individual sample sizes in zones.

Recreational walking did not vary significantly across census tracts or zones, in either two-level or cross-classified models.

Table 1. Results of unconditional two-level and cross-classified multilevel models predicting the number of 10-minute walking episodes for any motive, utilitarian purposes, and recreational reasons

Variance partitioning	Walking for any motive	Utilitarian walking	Recreational walking
Census tracts			
Variance component			
$\sigma^2_{individuals}$	3.12	5.56	5.31
$\tau_{census\ tracts}$	0.0003	0.099	0.004
χ^2 (df); p-value	142.40 (111); 0.024	157.27 (111); 0.003	123.69 (111); 0.194
% variation attributable to census tracts	0.01	1.75	0.08
Zones			
Variance component			
$\sigma^2_{individuals}$	3.12	5.13	5.29
τ_{zones}	0.0002	0.169	0.006
χ^2 (df); p-value	292.35 (269); 0.157	326.49 (269); 0.009	215.58 (269); $>$ 0.500
% variation attributable to zones	0.01	3.19	0.11
Cross-classified model			
Variance component			
$\sigma^2_{individuals}$	2.79	4.89	5.21
$\tau_{census\ tracts}$	0.031	0.068	0.006
χ^2 (df); p-value	142.10 (111); 0.025	140.52 (111); 0.030	123.29 (111); 0.200
τ_{zones}	0.016	0.121	0.005
χ^2 (df); p-value	273.21 (269); 0.417	294.23 (269); 0.139	215.23 (269); $>$ 0.500
% variation attributable to level-2 units (zones and census tracts)	1.66	3.72	0.21
% variation across level-2 units attributable to census tracts	65.96	35.98	54.55
% variation across level-2 units attributable to zones	34.04	64.02	45.45

Results of multilevel associations between area-level active living potential and walking when areas are operationalised by census tracts vs. zones

Results of two-level multilevel models for census tracts and zones are presented in Tables 2 and 3 respectively. Residents of census tracts categorised in the higher tertile of accessibility to services reported significantly ($p<0.05$) more 10-minute episodes of walking for any motive (event rate ratio [ERR]: ERR: 1.38; 95% Confidence interval [95%CI]: 1.12, 1.72) and for utilitarian purposes (ERR: 1.58; 95%CI: 1.11, 2.23). Adjusting models for census tract socioeconomic variable and individual characteristics marginally attenuated associations which remained statistically significant. Land use mix and population density were not associated with any type of walking. In final models (Model 3), a greater proportion of university graduates in the census tract was significantly associated with more walking episodes for any motive (ERR: 1.23; 95%CI: 1.04, 1.46) and for utilitarian purposes (ERR: 1.32; 95%CI: 1.01, 1.72).

Across zones (Table 3), residing in diverse and high accessibility central urban zones was associated with more episodes of walking for any motives and for utilitarian purposes. Specifically, in fully adjusted models and in comparison to residents of low-density suburban zones, residents of central urban zones with high accessibility reported more 10-minute episodes of walking for any motive (ERR: 1.48; 95%CI: 1.02, 2.14) and for utilitarian purposes (ERR: 1.98; 95%CI: 1.02, 3.86). Residents of diverse central urban zones were also more likely to cumulate more episodes of utilitarian walking (ERR: 1.93; 95%CI: 1.01, 3.71). Zone-level proportion of university graduate was only marginally associated with more episodes of walking for any motive (ERR: 1.16; 95%CI: 0.99, 1.37).

None of the census tract and zone level exposure variables was associated with recreational walking, for this reason results were not presented in the tables.

Table 2. Results of census tract models predicting the number of 10-minute walking episodes for any motive and utilitarian purposes as a function of active living and socioeconomic area exposures

	Walking for any motive			Utilitarian walking		
Census tracts	Model 1 a	Model 2 b	Model 3 c	Model 1 a	Model 2 b	Model 3 c
Fixed effects	ERR (95% CI) d	ERR (95% CI) d	ERR (95% CI) d	ERR (95% CI) d	ERR (95% CI) d	ERR (95% CI) d
Population density						
Lower tertile	1.00	1.00	1.00	1.00	1.00	1.00
Mid tertile	0.88 (0.72, 1.08)	0.92 (0.75, 1.13)	0.95 (0.77, 1.16)	0.91 (0.65, 1.26)	0.98 (0.70, 1.36)	1.01 (0.73, 1.39)
Higher tertile	0.88 (0.70, 1.10)	0.94 (0.75, 1.19)	0.98 (0.79, 1.24)	0.83 (0.58, 1.20)	0.93 (0.64, 1.35)	0.98 (0.68, 1.42)
Land use mix						
Lower tertile	1.00	1.00	1.00	1.00	1.00	1.00
Mid tertile	1.04 (0.87, 1.23)	1.05 (0.88, 1.25)	1.05 (0.88, 1.24)	1.05 (0.79, 1.39)	1.07 (0.81, 1.41)	1.06 (0.81, 1.40)
Higher tertile	1.09 (0.90, 1.31)	1.13 (0.93, 1.37)	1.15 (0.95, 1.39)	1.15 (0.85, 1.55)	1.20 (0.89, 1.62)	1.24 (0.93, 1.67)
Accessibility of services						
Lower tertile	1.00	1.00	1.00	1.00	1.00	1.00
Mid tertile	1.04 (0.85, 1.27)	1.04 (0.85, 1.27)	1.05 (0.87, 1.28)	1.03 (0.75, 1.43)	1.02 (0.73, 1.40)	1.05 (0.76, 1.44)
Higher tertile	1.38 (1.12, 1.72)*	1.33 (1.07, 1.65)*	1.32 (1.06, 1.63)*	1.58 (1.11, 2.23)*	1.46 (1.04, 2.07)*	1.46 (1.04, 2.05)*
Proportion of university graduates						
Lower tertile		0.96 (0.81, 1.14)	0.96 (0.82, 1.14)		0.96 (0.74, 1.26)	0.97 (0.75, 1.26)
Mid tertile		1.00	1.00		1.00	1.00
Higher tertile		1.20 (1.01, 1.41)*	1.23 (1.04, 1.46)*		1.28 (0.98, 1.67)†	1.32 (1.01, 1.72)*

[a] Model 1: Model with main exposure variables.
[b] Model 2: Model with main exposure and socioeconomic variables.
[c] Model 3: Model with main exposure and socioeconomic variables, adjusted for age, sex, education.
[d] ERR (95%CI): Event rate ratio and 95% confidence intervals.
[*] $p < 0.05$; [†] $p < 0.10$.

Table 3. Results of zones models predicting the number of 10-minute walking episodes for any motive and utilitarian purposes as a function of active living and socioeconomic area exposures

	Walking for any motive			Utilitarian walking		
Zones	Model 1 a	Model 2 b	Model 3 c	Model 1 a	Model 2 b	Model 3 c
Fixed effects	ERR (95% CI) d	ERR (95% CI) d	ERR (95% CI) d	ERR (95% CI) d	ERR (95% CI) d	ERR (95% CI) d
Categories of exposure in zones						
Low-density suburban	1.00	1.00	1.00	1.00	1.00	1.00
Middle-density suburban	1.03 (0.70, 1.52)	1.00 (0.68, 1.48)	0.98 (0.67, 1.44)	1.27 (0.64, 2.52)	1.24 (0.62, 2.45)	1.15 (0.58, 2.27)
Suburban / urban axial	0.95 (0.65, 1.39)	0.98 (0.67, 1.43)	1.01 (0.70, 1.47)	1.12 (0.57, 2.19)	1.17 (0.60, 2.28)	1.16 (0.60, 2.25)
Mix urban / suburban	1.02 (0.70, 1.48)	1.07 (0.73, 1.56)	1.09 (0.76, 1.58)	1.36 (0.70, 2.62)	1.43 (0.74, 2.77)	1.42 (0.74, 2.71)
Urban residential	0.91 (0.61, 1.36)	0.97 (0.64, 1.46)	0.99 (0.66, 1.48)	1.16 (0.58, 2.32)	1.23 (0.61, 2.47)	1.21 (0.61, 2.41)
Diverse central urban	1.19 (0.82, 1.72)	1.21 (0.83, 1.75)	1.25 (0.87, 1.80)	1.82 (0.94, 3.53)†	1.87 (0.97, 3.60)†	1.93 (1.01, 3.71)*
Central urban with high accessibility	1.40 (0.96, 2.04)†	1.43 (0.98, 2.09)†	1.48 (1.02, 2.14)*	1.99 (1.01, 3.92)*	1.99 (1.02, 3.89)*	1.98 (1.02, 3.86)*
Proportion of university graduates						
Lower tertile		1.01 (0.84, 1.21)	1.01 (0.85, 1.21)		1.07 (0.81, 1.43)	1.08 (0.81, 1.44)
Mid tertile		1.00	1.00		1.00	1.00
Higher tertile		1.16 (0.98, 1.37)†	1.16 (0.99, 1.37)†		1.23 (0.94, 1.61)	1.22 (0.93, 1.60)

[a] Model 1: Model with main exposure variables.
[b] Model 2: Model with main exposure and socioeconomic variables.
[c] Model 3: Model with main exposure and socioeconomic variables, adjusted for age, sex, education.
[d] ERR (95%CI): Event rate ratio and 95% confidence intervals.
[*] $p < 0.05$; [†] $p < 0.10$.

Assessing the relative influence of area-level active living potential and socioeconomic status on walking in cross-classified multilevel models

Results of cross-classified associations between active living potential and socioeconomic status and walking behaviours are presented in Table 4. Active living potential was significantly associated with walking for any, recreational, and utilitarian motives. Residents of central urban zones with high accessibility to services reported a weekly average of 33 10-minute episodes of walking for any motive per week (ERR: 1.42; 95%CI: 0.96, 2.11), in comparison to 23 episodes among residents of low-density suburban zones. People living in diverse and high accessibility central urban zones respectively cumulated on average 20 episodes (ERR: 1.80; 95%CI: 0.96, 3.38) and 22 episodes per week (ERR: 1.98; 95%CI: 1.03, 3.78) compared to 11 episodes in low-density suburban zones. In contrast, residents of diverse central urban zones were marginally less likely (p=0.057) to walk for recreational purposes in comparison to residents of low-density suburban zones (ERR: 0.66; 95%CI: 0.43, 1.01).

Adjusting models for the effect of the socioeconomic status of census tract (Model 2) and individual characteristics (Model 3) had marginal influence on the effects of the built environment on walking. In the final models, a greater proportion of university graduates in the census tract was positively associated with walking for any (ERR: 1.25; 95%CI: 1.06, 1.49), utilitarian (ERR: 1.33; 95%CI: 1.04, 1.71), and recreational (ERR: 1.24; 95%CI: 1.01, 1.54) motives.

Table 4. Results of cross-classified multilevel models predicting the number of 10-minute walking episodes for any motive, utilitarian purposes, and recreational reasons as a function of active living and socioeconomic area exposures

	Model 1 [a]	Model 2 [b]	Model 3 [c]
Walking for any motive	**ERR (95% CI) [d]**	**ERR (95% CI) [d]**	**ERR (95% CI) [d]**
Categories of exposure in zones			
Low-density suburban	1.00	1.00	1.00
Middle-density suburban	1.03 (0.69, 1.53)	0.98 (0.66, 1.46)	0.97 (0.65, 1.44)
Suburban / urban axial	0.94 (0.64, 1.39)	0.93 (0.63, 1.36)	0.95 (0.65, 1.39)
Mix urban / suburban	1.03 (0.71, 1.52)	1.04 (0.71, 1.52)	1.07 (0.73, 1.55)
Urban residential	0.93 (0.61, 1.40)	0.95 (0.63, 1.43)	0.97 (0.65, 1.46)
Diverse central urban	1.19 (0.81, 1.75)	1.16 (0.80, 1.70)	1.22 (0.84, 1.78)
Central urban with high accessibility	1.42 (0.96, 2.11)†	1.37 (0.93, 2.02)	1.42 (0.97, 2.09)†
Proportion of university graduates in census tracts			
Lower tertile		1.01 (0.84, 1.20)	1.02 (0.86, 1.21)
Middle tertile		1.00	1.00
Higher tertile		1.23 (1.04, 1.46)*	1.25 (1.06, 1.49)*
Utilitarian walking	**ERR (95% CI) [d]**	**ERR (95% CI) [d]**	**ERR (95% CI) [d]**
Categories of exposure in zones			
Low-density suburban	1.00	1.00	1.00
Middle-density suburban	1.27 (0.66, 2.45)	1.23 (0.65, 2.36)	1.16 (0.61, 2.22)
Suburban / urban axial	1.14 (0.60, 2.16)	1.14 (0.61, 2.14)	1.14 (0.61, 2.13)
Mix urban / suburban	1.35 (0.72, 2.54)	1.38 (0.75, 2.58)	1.39 (0.75, 2.57)
Urban residential	1.17 (0.60, 2.28)	1.21 (0.63, 2.35)	1.20 (0.62, 2.31)
Diverse central urban	1.80 (0.96, 3.38)†	1.77 (0.95, 3.30)†	1.87 (1.00, 3.47)*
Central urban with high accessibility	1.98 (1.03, 3.78)*	1.91 (1.01, 3.61)*	1.92 (1.02, 3.65)*
Proportion of university graduates in census tracts			
Lower tertile		1.03 (0.80, 1.34)	1.06 (0.82, 1.36)
Middle tertile		1.00	1.00
Higher tertile		1.30 (1.01, 1.68)*	1.33 (1.04, 1.71)*
Recreational walking	**ERR (95% CI) [d]**	**ERR (95% CI) [d]**	**ERR (95% CI) [d]**
Categories of exposure in zones			
Low-density suburban	1.00	1.00	1.00
Middle-density suburban	0.78 (0.51, 1.21)	0.73 (0.47, 1.13)	0.75 (0.49, 1.16)
Suburban / urban axial	0.77 (0.50, 1.17)	0.74 (0.49, 1.13)	0.75 (0.50, 1.13)
Mix urban / suburban	0.76 (0.50, 1.15)	0.76 (0.50, 1.15)	0.78 (0.52, 1.17)
Urban residential	0.70 (0.44, 1.11)	0.72 (0.45, 1.14)	0.76 (0.49, 1.19)
Diverse central urban	0.66 (0.43, 1.01)†	0.64 (0.41, 0.98)*	0.65 (0.43, 0.99)*
Central urban with high accessibility	0.86 (0.55, 1.33)	0.82 (0.53, 1.27)	0.86 (0.56, 1.31)
Proportion of university graduates in census tracts			
Lower tertile		0.95 (0.77, 1.19)	0.95 (0.77, 1.18)
Middle tertile		1.00	1.00
Higher tertile		1.20 (0.97, 1.49)†	1.24 (1.01, 1.54)*

[a] Model 1: Model with main exposure variables; [b] Model 2: Model with main exposure and socioeconomic variables;
[c] Model 3: Model with main exposure and socioeconomic variables, adjusted for age, sex, education.
[d] ERR (95%CI): Event rate ratio and 95% confidence intervals. * $p < 0.05$; † $p < 0.10$.

DISCUSSION

In order to examine the contribution of designing homogenous spatial units along the exposure of interest, the objectives of this study were: 1) to examine and compare the predictive value of active living potential operationalised at the census tract and zone levels; and 2) in cross-classified multilevel models, to assess the relative influence of active living potential and socioeconomic exposure variables on walking behaviours using different area units that are homogeneous along these exposures, respectively zones and census tracts. These objectives were pursued in light of previous research showing that the construct validity of census tracts for delimiting areas in relation to active living potential exposures variable may be limited, therefore threatening the internal validity of studies examining area effects on walking at this level of analysis. By design zones homogeneous along active living potential and examining associations between this exposure and walking behaviour at this level, the internal validity of the study is increased given the limited biases related to exposure misclassification and association misestimation.

As in other studies, greater geographical accessibility to proximity services was an important determinant of walking (De Bourdeaudhuij et al., 2003; Giles-Corti & Donovan, 2002b; Humpel, Owen, Leslie, Marshall, Bauman, & Sallis, 2004; Krizek & Johnson, 2006; Tilt, Unfried, & Roka, 2007). Greater accessibility at the census tract level was associated with more episodes of walking for any motive and for utilitarian purposes. Zones characterised by the highest values of accessibility to proximity services in conjunction with higher population density (high accessibility central urban zones) and greater land use mix (diverse central urban zones) were associated with greater involvement in walking for any motive and for utilitarian purposes.

Although the magnitude of the association between active living potential and walking was greater in the zone model, estimates were similar to those obtained in the census tract models. This was expected given the relative homogeneity of census tracts along accessibility to services. Nonetheless, using area units homogenous along exposures increases the internal validity of the study design and therefore improves the soundness of statistical inference.

In census tract models, population density and land use mix (more heterogeneous within these spatial units) were not associated with walking. Results provide some support for previous research reporting associations between residential density and different types of walking (utilitarian vs. recreational), but not with the amount of walking (Forsyth, Oakes, Schmitz, & Hearst, 2007). In the current study, the measurement of land use mix was dependent on the size of dissemination areas which are defined in part by a population threshold; they are likely to span a larger territory and therefore encompass more land uses when located in suburban areas as population density is lower in these types of areas. Although this was not a problem for designing zones given that dissemination areas were classified based on their statistical similarity and spatial contiguity, in studies wherein land use mix is used as a predictor of walking behaviours, it should be measured at scales others than those defined by a population threshold (Frank et al., 2005).

The cross-classified multilevel model allowed for the estimation of the distinct influence of active living potential and socioeconomic status on walking behaviours. Results showed that when exposure is operationalised by appropriate area units, i.e. zones homogeneous along active living potential and census tracts homogeneous along socioeconomic status, both exposures were significantly associated with walking for any, utilitarian, and recreational motives. These results reveal that the different dimensions of residential areas influencing walking behaviours operate at diverse spatial scales which may be overlapping rather than hierarchically nested within one another.

Finally, the results draw attention to the specificity of certain environmental contexts for influencing walking. Zones were designed based on attributes of the built environment potentially supporting greater involvement in utilitarian walking; these attributes may not be relevant for fostering greater recreational walking in the population. Indeed, findings showed that diverse central urban zones conducive to more episodes of utilitarian walking were less conducive to recreational walking than low-density suburban zones. This suggests that specific environmental features may be associated with different motives for walking. In addition, recreational walking may be influenced by other dimensions of environments not examined in this study, such as accessibility to public open spaces and parks, safety from traffic and crime, and aesthetics of the surroundings (Giles-Corti et al., 2005; Li et al., 2005; Owen et al., 2004; Sallis, Cervero, Ascher, Henderson, Kraft, & Kerr, 2006).

Increasing non-motorised geographical accessibility of proximity services should not be seen as the driving force of public health policies aiming at increasing active living in the population and reducing trends in overweight and obesity. Features of the built environment are often correlated to create distinct area contexts, as was illustrated by the seven "categories of exposure". Rather, what seems to emerge from the zone-level model is the significance of density: density of destinations accessible by foot, density of population, and mixed land uses. Future natural experiment studies are needed to explore the potential causal role of increasing densities in the community and level of walking.

Strengths and limitations

Zones were delimited for methodological and aetiological purposes with the aim of limiting measurement errors and increasing internal validity of the study design. For this, designing homogeneous zones limits the external validity of the study, where generalisability to other contexts is limited, for example residential neighbourhoods and spaces of action of public health which are not likely to be homogeneous with respect to exposures.

Results are subject to limitations due to sampling procedures and statistical power. A representative sample of respondents was randomly selected within the 112 census tracts. Zones were designed independently of this sampling. Small individual sample sizes in zones may influence estimation of effect sizes. As this study was a pilot investigation to assess the feasibility and added-value of designing homogeneous zones for studying environmental determinants of health, future studies should endeavour to define homogeneous zones a priori, and then randomly sample participants within homogeneous and contrasting zones. The cross-sectional design of the study prevents from ascertaining whether characteristics of the built and social environments were the catalyst for walking or whether persons involved in greater amounts of walking chose to live in environments with greater active living potential. Longitudinal studies, wherein change over time in both individual walking behaviours and environments, are warranted. Finally, another limitation of the study has to do with self-reported minutes of walking and disentangling recreational and utilitarian walking.

CONCLUSION

Findings of the study suggest that designing spatial units maximising homogeneity of exposures is advantageous to defining areas as it strengthens internal validity without becoming unwieldy. Replication studies are required to further assess the value of the novel method for designing zones relevant to other health indicators. For public health, the relevance of homogeneous zones lies in their design which allows for the optimal identification of environmental determinants of a given health indicator in order to inform interventions aiming at modifying areas to create supportive environments for health. It also draws attention to the need to inventory and investigate different area attributes potentially influencing health as part of the surveillance function of public health.

References

Cervero, R. (1996). Mixed land-uses and commuting: Evidence from the American housing survey. *Transportation Research Part a-Policy and Practice,* 30(5), 361-377.

Chaix, B., Leyland, A. H., Sabel, C. E., Chauvin, P., Rastam, L., Kristersson, H., et al. (2006). Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmö, Sweden, in 2001. *Journal of Epidemiology and Community Health,* 60, 427-435.

Chaix, B., Rosvall, M., Lynch, J., & Merlo, J. (2006). Disentangling contextual effects on cause specific mortality in a longitudinal 23-year follow-up study: impact of population density or socioeconomic environment? . *International Journal of Epidemiology,* 35, 633-643.

Cockings, S., & Martin, D. (2005). Zone design for environment and health studies using pre-aggregated data. *Social Science & Medicine,* 60(12), 2729-2742.

Craig, C. L., Marshall, A. L., Sjostrom, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., et al. (2003). International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc,* 35, 1381-1395.

Cummins, S., Macintyre, S., Davidson, S., & Ellaway, A. (2005). Measuring neighbourhood social and material context: generation and interpretation of ecological data from routine and non-routine sources. *Health & Place,* 11(3), 249-260.

De Bourdeaudhuij, I., Sallis, J. F., & Saelens, B. E. (2003). Environmental correlates of physical activity in a sample of Belgian adults. *American Journal of Health Promotion,* 18(1), 83-92.

Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health,* 91(11), 1783-1789.

ESRI. (2006). ArcGIS 9.2. Redlands, CA.

Fisher, K., Li, F., & Michael, Y. (2004). Neighborhood-level influences on physical activity among older adults: a multi-level analysis. . *Journal of Aging and Physical Activity,* 12, 45-63.

Forsyth, A., Oakes, J. M., Schmitz, K. H., & Hearst, M. (2007). Does residential density increase walking and other physical activity? *Urban Studies,* 44(4), 679-697.

Frank, L. D., & Pivo, G. (1995). Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking. *Transportation Research Record,* 1466, 44-52.

Frank, L. D., Schmid, T. L., Sallis, J. F., Chapman, J., & Saelens, B. E. (2005). Linking objectively measured physical activity with objectively measured urban form - Findings from SMARTRAQ. *American Journal of Preventive Medicine,* 28(2), 117-125.

Frohlich, K. L., Potvin, L., Chabot, P., & Corin, E. (2002). A theoretical and empirical analysis of context: neighbourhoods, smoking and youth. *Social Science & Medicine,* 54, 1401-1417.

Galea, S., & Ahern, J. (2006). Invited commentary: Considerations about specificity of associations, causal pathways, and heterogeneity in multilevel thinking. *American Journal of Epidemiology,* 163(12), 1079-1082.

Gauvin, L., Richard, L., Craig, C. L., Spivock, M., Riva, M., Forster, M., et al. (2005). From walkability to active living potential - An "ecometric" validation study. *American Journal of Preventive Medicine,* 28(2), 126-133.

Gauvin, L., Riva, M., Barnett, T. A., Richard, L., Craig, C. L., Spivock, M., et al. (2008). Association between Neighborhood Active Living Potential and Walking. *American Journal of Epidemiology*, 167(8), 944-953.

Gauvin, L., Robitaille, E., Riva, M., McLaren, L., Dassa, C., & Potvin, L. (2007). Conceptualizing and operationalizing neighbourhoods: the conundrum of identifying territorial units. *Canadian Journal of Public Health,* 98(suppl 1), S18-S26.

Giles-Corti, B., Broomhall, M. H., Knuiman, M., & al., e. (2005). Increasing walking: how important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine,* 28(Suppl 2), 169-176.

Giles-Corti, B., & Donovan, R. J. (2002a). The relative influence of individual, social and physical environment determinants of physical activity. *Social Science & Medicine,* 54(12), 1793-1812.

Giles-Corti, B., & Donovan, R. J. (2002b). Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. *Preventive Medicine,* 35(6), 601-611.

Haynes, R., Daras, K., Reading, R., & Jones, A. (2007). Modifiable neighbourhood units, zone design and residents ' perceptions. *Health & Place,* 13(4), 812-825.

Humpel, N., Owen, N., Leslie, E., Marshall, A., & Bauman, A. (2004). Associations of location and perceived environmental attributes with walking in neighbourhoods. *American Journal of Health Promotion,* 18, 239-242.

Humpel, N., Owen, N., Leslie, E., Marshall, A. L., Bauman, A. E., & Sallis, J. F. (2004). Associations of location and perceived environmental attributes with walking in neighborhoods. *American Journal of Health Promotion,* 18(3), 239-242.

King, W. C., Brach, J. S., Belle, S., Killingsworth, R., Fenton, M., & Kriska, A. M. (2003). The relationship between convenience of destinations and walking levels in older women. *American Journal of Health Promotion,* 18, 74-82.

Krizek, K. J., & Johnson, P. J. (2006). Proximity to trails and retail: Effects on urban cycling and walking. *Journal of the American planning association,* 72(1), 33-42.

Lee, C., & Vernez Moudon, A. (2006). Correlates of Walking for Transportation or Recreation Purposes. *Journal of Physical Activity and health,* 3 (Suppl 1), S77.

Lee, I., & Paffenbarger, R. J. (2000). Association of light, moderate, and vigorous intensity physical activity with longevity: The Harvard Alumni Study. *American Journal of Epidemiology,* 151, 293-299.

Li, F. Z., Fisher, K. J., Brownson, R. C., & Bosworth, M. (2005). Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *Journal of Epidemiology and Community Health,* 59(7), 558-564.

Oakes, J. M., & Rossi, P. H. (2003). The measurement of SES in health research: current practice and steps toward a new approach. *Social Science & Medicine,* 56(4), 769-784.

Osypuk, T. L., & Galea, S. (2007). What level macro? Choosing appropriate level to assess how place influences population health. In S. Galea (Ed.), Macrosocial determinants of population health pp. 399-438). New York: Springer.

Owen, N., Humpel, N., Leslie, E., Bauman, A., & Sallis, J. F. (2004). Understanding environmental influences on walking - Review and research agenda. *American Journal of Preventive Medicine,* 27(1), 67-76.

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., et al. (1995). Physical activity and public health - a recommendation from the Centers for disease control and prevention and the American college of sports medicine. *Jama-Journal of the American Medical Association,* 273(5), 402-407.

Pearce, J., Witten, K., & Bartie, P. (2006). Neighbourhoods and health: A GIS approach to measuring community resource accessibility. *Journal of Epidemiology and Community Health,* 60(5), 389-395.

Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health,* 55(2), 111-122.

Pikora, T. J., Giles-Corti, B., Knuiman, M. W., Bull, F. C., Jamrozik, K., & Donovan, R. J. (2006). Neighborhood environmental factors correlated with walking near home: Using SPACES. *Medicine and Science in Sports and Exercise,* 38(4), 708-714.

Popay, J., Thomas, C., Williams, G., Bennett, S., Gattrell, A., & Bostock, L. (2003). A proper place to live: Health inequalities, agency and the normative dimensions of space. *Social Science & Medicine,* 57(1), 55-69.

Raudenbush, S., & Bryk, A. (2002). Hierarchical linear models: applications and data analysis methods Newbury Park, CA: Sage.

Raudenbush, S., Bryk, A., & Congdon, R. (2005). HLM: Hierarchical Linear and Nonlinear Modeling (Version 6.04) Chicago: Scientific Software International.

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology*, 29, 1-41.

Riva, M., Gauvin, L., & Barnett, T. A. (2007). Toward the next generation of research into small area effects on health: A synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology and Community Health*, 61(10), 853-861.

Riva, M., Apparicio, P., Gauvin, L., & Brodeur, J-M. (In Press). Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for designing optimal zones. *International Journal of Health Geographic.*

Ross, N. A., Tremblay, S., & Graham, K. (2004). Neighbourhood influences on health in Montreal, Canada. *Social Science & Medicine*, 59(7), 1485-1494.

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine*, 25(2), 80-91.

Sallis, J. E., Cervero, R. B., Ascher, W., Henderson, K. A., Kraft, M. K., & Kerr, J. (2006). An ecological approach to creating active living communities. *Annual Review of Public Health*, 27, 297-322.

Siegel, P. Z., Brackbill, R. M., & Heath, G. W. (1995). The epidemiology of walking for exercise: implications for promoting activity among sedentary groups. *American Journal of Public Health*, 85, 706-710.

Statistics Canada. 2001 Census preview of products and Services. Ottawa: Statistics Canada.

Statistics Canada Census Operations Division. (2003). 2001 Census Dictionary. Minister of Industry, Canada.

Theil, H. (1972). Statistical decomposition analysis. Amsterdam: North-Holland.

Theil, H., & Finezza, A. (1971). A note on the measurement of racial integration of schools by means of informational concepts. *Journal of Mathematical Sociology*, 1, 187-194.

Tilt, J. H., Unfried, T. M., & Roka, B. (2007). Using objective and subjective measures of neighborhood greenness and accessible destinations for understanding walking trips and BMI in Seattle, Washington. *American Journal of Health Promotion,* 21(4), Suppl. 371-379.

Transportation Research Board and Institute of Medicine of the National Academies (TRB & IOM). (2005). Does the built environment influence physical activity? Examining the evidence. Special Report 282. Transportation Research Board. Washington, DC.

United States Department of Health and Human Services (USDHHS). (1996). Physical Activity and health: A report from the Surgeon General. Atlanta, GA Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

USDHHS. (1996). Physical Activity and health: A report from the Surgeon General. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

World Cancer Research Fund / American Institute for Cancer Research (WCRF / AICR). (2007). Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR.

DISCUSSION

Par un rappel des principaux résultats de la thèse, rappel qui est structuré selon les deux objectifs spécifiques de recherche, ce dernier chapitre propose une prise de position sur la question de recherche, c.-à-d. l'apport de la création d'unités spatiales homogènes au niveau de l'exposition d'intérêt pour formuler des inférences sur les effets de milieux sur la santé. Cette discussion mène à la proposition d'un schéma décisionnel pour guider l'opérationnalisation du milieu. Par la suite, les forces et les limites du projet de doctorat sont mises en relief, des avenues pour des recherches futures sont proposées et la portée des résultats pour la recherche et l'intervention en santé publique et en promotion de la santé est discutée.

Synthèse et signification des résultats

Homogénéité des secteurs de recensement et création des zones

Le deuxième article a permis d'évaluer le degré d'homogénéité des secteurs de recensement au niveau 1) des trois indicateurs du potentiel de vie active et 2) des catégories d'exposition correspondant à des potentiels de vie active différents. Dans un premier temps, les résultats indiquent que, alors que les secteurs sont homogènes en termes d'accessibilité aux services (la variabilité inter-secteur est de beaucoup supérieure à la variabilité intra-secteur), ils sont plutôt hétérogènes en fonction de la densité de la population et de la mixité des occupations du sol. Ensuite, il a été observé qu'une majorité des secteurs de recensement (55.5%) est caractérisée par une combinaison de trois catégories d'exposition ou plus, ces catégories étant substantivement différentes entre elles au niveau de leur potentiel de vie active. Ainsi, parce qu'ils comprennent des environnements qui sont à la fois plus favorables pour la marche et d'autres qui le sont moins, plus de la moitié des secteurs de recensements sont hétérogènes en termes de potentiel de vie active; ils sont donc peu appropriés pour opérationnaliser les milieux pour l'étude des déterminants environnementaux de la marche. En contrepartie, les secteurs de recensement sont plus homogènes en regard des caractéristiques socioéconomiques, notamment au niveau de la proportion de personnes ayant un diplôme universitaire. Toutefois, la distribution spatiale de ces variables est dissemblable de celle du potentiel de vie active, ce qui indique que des unités spatiales différentes puissent optimiser chacune de ces expositions.

L'approche développée pour créer des zones optimales homogènes au niveau d'expositions est conviviale tant dans sa conceptualisation que dans son application. Elle est basée sur la classification des aires de diffusion ayant un potentiel de vie active semblable, puis de leur regroupement dans l'espace en fonction de leur similarité et contiguïté. Pour cela, l'approche requiert une connaissance de base des méthodes de classification statistique et des applications simples disponibles dans les systèmes d'information géographique. En outre, l'application de cette approche permet de maximiser avec succès l'homogénéité des expositions dans les unités et de maximiser le plus possible les différences entre elles.

Effets de milieux sur la marche : comparaison des résultats lorsque le milieu est opérationnalisé par les zones plutôt que par les secteurs de recensement

Le troisième article de la thèse a examiné l'influence du potentiel de vie active des milieux sur la pratique de la marche, en comparant les résultats obtenus lorsque le milieu est opérationnalisé par les nouvelles unités spatiales plutôt que par les secteurs de recensement.

Les analyses ont d'abord été menées pour les individus imbriqués dans les secteurs de recensement et pour les individus imbriqués dans les zones. Tel qu'il a été observé dans d'autres études (Cerin et al., 2007; De Bourdeaudhuij et al., 2003; Giles-Corti & Donovan, 2002b; Humpel et al., 2004; Krizek & Johnson, 2006; Tilt, Unfried, &

Roka, 2007; Lovasi et al., 2008), les résultats démontrent que l'accessibilité géographique aux services de proximité apparaît comme un déterminant important de la marche. Une plus grande accessibilité est associée à des épisodes de marche totale et de marche utilitaire plus nombreux au niveau des secteurs de recensement. Les zones caractérisées par les valeurs d'accessibilité les plus élevées et jumelées à une forte densité de population (*central urban area with high accessibility*) et à une mixité d'occupation du sol importante (*diverse central urban*), sont associées à un nombre plus élevé d'épisodes de marche totale et de marche utilitaire. Bien que la taille des coefficients soit plus importante dans les zones, les résultats sont plutôt similaires à ceux obtenus au niveau des secteurs de recensement (tel qu'indiqué par le chevauchement des intervalles de confiance). Ces résultats ne sont pas étonnants étant donné l'homogénéité des secteurs de recensement en termes d'accessibilité géographique aux services. Néanmoins, les inférences quant à l'influence du potentiel de vie active sur les habitudes de marche sont plus justes au niveau des zones car les erreurs de mesure sont limitées. Par contre, la densité de population et la mixité de l'occupation du sol, plus homogènes au niveau des zones, agissent de concert avec l'accessibilité pour configurer des environnements ayant un plus grand potentiel de vie active. De telles associations ne sont pas observées au niveau des secteurs de recensement.

Les résultats des modèles multiniveaux avec classification croisée démontrent que, lorsque le potentiel de vie active et le statut socioéconomique sont opérationnalisés par des unités spatiales appropriées c.-à-d. respectivement les zones et les secteurs de recensement, ces expositions sont associées significativement avec la marche utilitaire, récréative et la marche pour tous les motifs (totale). Ces résultats font donc état que les différentes dimensions de l'environnement qui influencent la marche opèrent à des échelles spatiales différentes, qui se chevauchent plutôt qu'elles ne s'imbriquent.

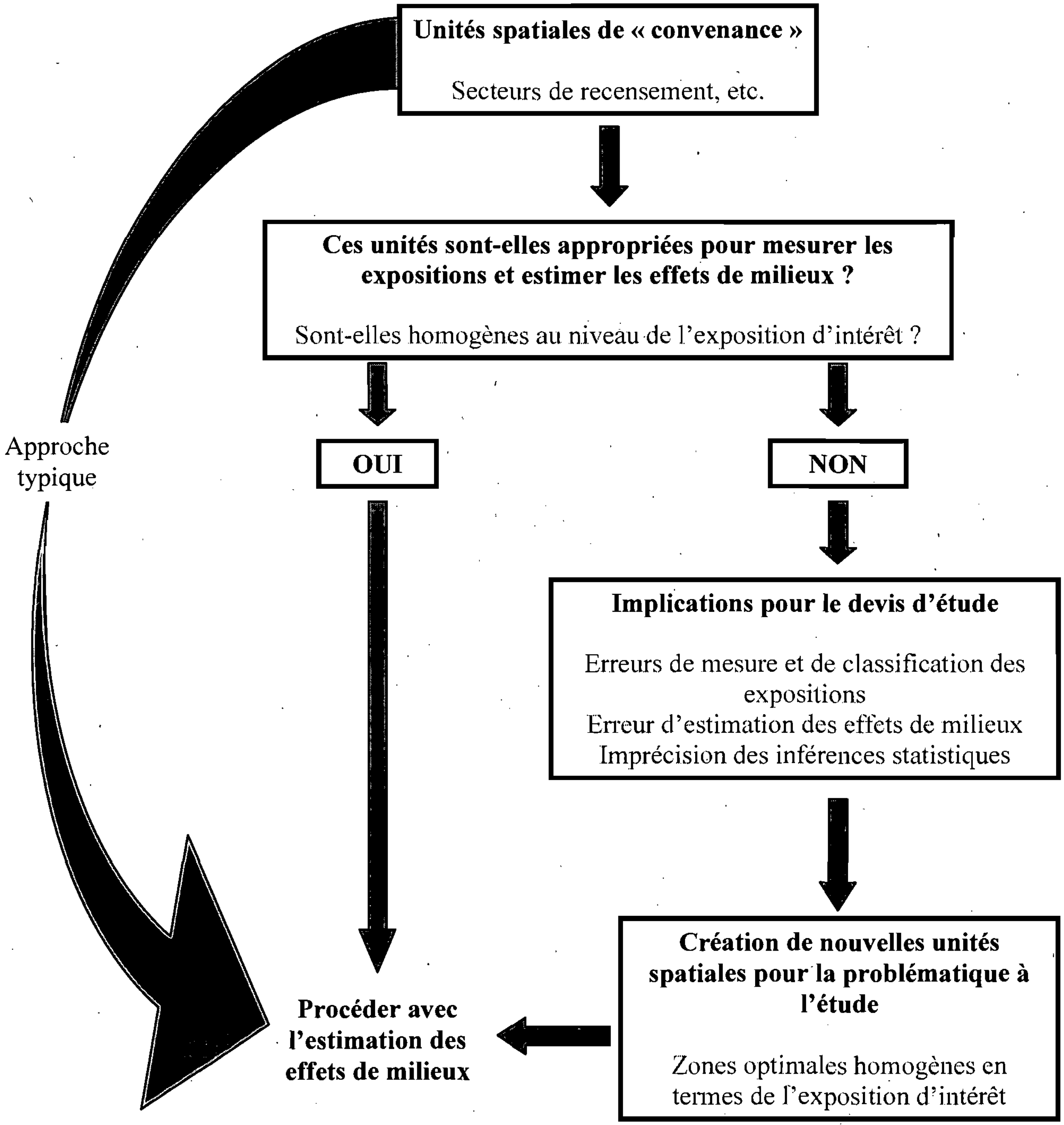
Apport du développement et de l'application d'une approche pour créer des unités spatiales pour la compréhension des effets de milieux sur la santé

Établir l'homogénéité des expositions dans l'unité spatiale d'analyse permet d'évaluer les erreurs de mesure des expositions possibles pouvant mener à une mauvaise estimation (*misestimation*) des effets de milieux. Or, jusqu'à présent, cette considération méthodologique a été peu abordée dans l'étude des effets de milieux sur la santé. À la lumière des résultats des articles empiriques, il est proposé que le schéma décisionnel présenté à la Figure 1 de l'article 2 (version anglaise, p. 78) serve à orienter l'opérationnalisation du milieu. Ce schéma est présenté en version française à la page suivante.

Dans un premier temps, il s'agit de déterminer si l'unité spatiale d'analyse disponible, c.-à-d. les unités spatiales de « convenance », est homogène en regard à l'exposition d'intérêt avant de procéder aux analyses pour mesurer les effets de milieux. À ce titre, il est possible que des unités administratives facilement accessibles soient valides pour la problématique à l'étude. Par exemple, dans la thèse, si la question de recherche avait porté sur l'influence unique de l'accessibilité aux services du milieu de résidence sur la marche, ou encore sur l'influence du statut socioéconomique mesuré par le niveau d'éducation, l'utilisation des secteurs de recensement aurait été appropriée.

Mais étant donné l'hétérogénéité des secteurs de recensement en regard aux autres variables d'exposition et aux catégories d'exposition, l'approche qui consiste à créer des unités spatiales qui maximisent l'homogénéité des expositions s'avère avantageuse pour définir les milieux, notamment par le renforcement de la validité interne qu'elle génère. La contribution de cette approche est d'autant plus intéressante qu'elle est relativement simple à appliquer et à répliquer. La classification statistique est une méthode statistique descriptive et exploratoire (plutôt qu'inférentielle) largement utilisée dans les sciences sociales pour regrouper

Figure 2. Schéma décisionnel pour opérationnaliser le milieu



différentes variables en classes distinctes, et ainsi créer des typologies (Harris, Sleight, & Webber, 2005). La méthode de classification employée, le *K-means*, est disponible dans une majorité de logiciels de traitement de données (par exemple, SPSS, SAS, STATA). La composante plus « complexe » de l'approche réside dans l'utilisation de systèmes d'information géographique. Or, l'utilité de ces derniers pour l'étude des déterminants de la santé est désormais reconnue et leur utilisation est beaucoup plus répandue.

Ainsi, par l'utilisation d'unités spatiales homogènes en termes d'exposition, la validité interne du devis de recherche sera accrue et la précision des inférences statistiques en sera augmentée. Néanmoins, il serait hâtif à ce stade-ci de statuer formellement sur la plus-value de l'utilisation d'unités homogènes. Pour ce faire, d'autres études sont requises pour explorer l'impact du choix d'autres indicateurs de santé et d'autres expositions pour la création de zones ayant une signification étiologique pour étudier d'autres états et indicateurs de santé. Par exemple, des variables d'exposition liées au potentiel de vie active des milieux pourraient être associées à des variables d'accessibilité à des aliments sains et à des services de restauration rapide, pour créer des unités spatiales appropriées pour comprendre les déterminants environnementaux du surpoids et de l'obésité. Qui plus est, la stratégie d'échantillonnage devra être cohérente avec la structure spatiale à l'étude. Dans la thèse, les données individuelles provenaient d'une étude dans laquelle les répondants ont été sélectionnés aléatoirement dans 112 secteurs de recensement sur l'Île de Montréal (Gauvin et al., 2008). Les zones ont été créées indépendamment de cet échantillonnage. Puisque la stratégie d'échantillonnage n'était pas stratifiée, la taille des effectifs individuels dans certaines zones est faible, ce qui peut expliquer que les résultats n'aient pas toujours atteint les seuils de signification conventionnels de $p < 0.05$. Les zones homogènes devraient être créées *a priori* de façon à structurer la stratégie d'échantillonnage.

Finalement, informé par les distributions spatiales différentes des facteurs d'exposition liés au potentiel de vie active et au statut socioéconomique, l'emploi de modèles multiniveaux avec classification croisée démontre que les niveaux écologiques d'influence peuvent se chevaucher, plutôt que s'imbriquer hiérarchiquement. Ces modèles sont donc plus cohérents avec la structure des influences écologiques sur la marche. Lorsque l'effet de deux expositions différentes sur un indicateur de santé est étudié, l'emploi de ces modèles limite les problèmes reliés à la multicolinéarité de la mesure des expositions, qui est « forcée » dans les devis typiques puisque les expositions sont mesurées dans une même unité spatiale. De plus, l'utilisation d'unités spatiales appropriées limite les enjeux soulevés par la problématique des unités géographiques modifiables.

Forces et limites du projet de doctorat et avenues de recherches futures

Ce projet de doctorat comporte certaines forces et limites. Sans aucun doute, la plus grande force réside dans la création d'unités spatiales homogènes au niveau de l'exposition d'intérêt. Cet effort s'inscrit dans la lignée de quelques études portant sur la question de la délimitation des frontières des milieux pour la recherche en santé (Cockings & Martin, 2005; Haynes et al., 2007; Haynes et al., 2008; Law et al., 2005; Martin et al., 2001; Nakaya, 2000; Openshaw, 1996; Openshaw & Rao, 1995; Reading et al., 1999; Tatalovich et al., 2006; The Project on Human Development in Chicago Neighborhoods http://www.icpsr.umich.edu/Phdcn). L'utilisation d'autres méthodes (bien qu'impliquant des recettes plus complexes) pour créer des unités spatiales est à investiguer; notamment, les méthodes de classifications statistiques floues (*fuzzy K-means*) (Duda, Hart, & Stork, 2001; Everitt, Landau, & Leese, 2001; Fortin & Dale, 2005), le design automatique de zones (Cockings & Martin, 2005; Haynes et al., 2007; Haynes et al., 2008; Martin et al., 2001; Nakaya, 2000; Openshaw, 1996; Openshaw & Rao, 1995) et les méthodes de classification spatiale (Lawson & Denison, 2002; Murray, 1999; Murray & Estivill-Castro, 1998) permettraient de bonifier l'opérationnalisation du milieu.

Les possibilités d'erreurs liées à une mauvaise classification des zones en termes de catégories d'exposition (*exposure misclassification*) sont limitées par la configuration des zones délimitées selon la distribution spatiale des variables d'exposition. Par contre, il est à noter que la mesure de mixité de l'occupation du sol, mesurée au niveau des aires de diffusion, comporte une certaine limite. Alors qu'il serait attendu que l'on observe une plus grande mixité du sol dans les milieux plus urbains, l'inverse s'est avéré : la mixité est plus importante en banlieue en raison de la plus grande superficie couverte par les aires de diffusion dans ces milieux. Pour la classification statistique et la création des zones, cela a eu peu d'effets puisque les variables ont été agrégées selon leur proximité statistique et spatiale : les aires de diffusion en milieux urbains présentaient un profil d'occupation du sol similaire entre elles; et inversement, le profil d'occupation du sol des aires de diffusion localisées en banlieue était similaire. Par contre, dans les recherches examinant l'occupation du sol comme variable de prédiction de la marche, cette variable devrait être mesurée à l'intérieur d'une unité géographique dont la taille n'est pas délimitée selon un seuil de population (i.e. aires de diffusion, secteurs de recensement) telles des zones de proximité (*buffer*) délimitées par un rayon de *n* mètres autour d'un centroïde prédéfini.

Les zones ont été créées pour identifier quels aspects de l'environnement bâti, mesurés de façon optimale, sont associés à la marche. La délimitation de zones homogènes est donc pour des fins méthodologiques et étiologiques dans un souci de limiter les erreurs de mesures et d'augmenter la validité interne des devis de recherche. Pour cela, la délimitation de zones homogènes en termes d'expositions mène à un problème de validité externe. En effet, la validité externe des résultats à d'autres milieux est limitée puisque les quartiers résidentiels dans d'autres environnements urbains ne sont peut-être pas comparables. Nonobstant, le volet généralisable de cette thèse est la méthode pour créer des zones, et non pas les zones comme telles.

En parallèle, il est peu probable qu'il y ait correspondance entre les zones et les perceptions que les résidents entretiennent des frontières de leur quartier. Bien que plusieurs personnes conçoivent habiter dans un quartier, il semble que les perceptions des frontières varient entre les résidents d'un même quartier (Coulton et al., 2001). Opérationnaliser les milieux en fonction des perceptions importe lorsque la notion de territoire partagé est liée à l'exposition d'intérêt, par exemple le capital social ou l'efficacité collective (Diez Roux, 2001; Osypuk & Galea, 2007). Cette notion n'est pas évoquée par le potentiel de vie active.

La nature transversale du devis de recherche pour étudier les déterminants environnementaux de la marche pose certaines limites. D'abord, il est impossible de déterminer la causalité des associations, c.-à-d. si les caractéristiques des milieux catalysent les comportements de marche. Des devis longitudinaux, dans lesquels les modifications des environnements et des comportements de marche sont suivies et mesurées à différents intervalles de temps permettraient de contrer ces limites. En ce sens, il serait souhaitable de tirer profit d'expériences dites « naturelles » (Cummins, Petticrew, Higgins, Findlay, & Sparks, 2005b; Petticrew, Cummins, Ferrell, Findlay, Higgins, Hoy et al., 2005), par exemple la création d'une voie piétonnière ou une densification des services de proximité, et d'évaluer leurs impacts sur la santé des populations locales affectées par ces développements. Ensuite, il est probable qu'un biais de sélection teinte la lecture des résultats car le devis ne permet pas de statuer si des personnes pratiquant régulièrement la marche choisissent de vivre dans des milieux qui en facilitent la pratique (Frank, Saelens, Powell, & Chapman, 2007; Handy, Cao, & Mokhtarian, 2006).

Parallèlement, il est important d'examiner plus en détail le temps d'exposition quotidien au milieu résidentiel. En effet, les activités des individus ont lieu dans différents espaces, que ce soit le lieu de travail ou les lieux de socialisation, tous ayant une influence potentielle sur la pratique de la marche. La compilation quotidienne des déplacements et la localisation spatiale des lieux d'origine et de destinations permettraient d'opérationnaliser les différents espaces d'activité des individus, les facteurs auxquels ils sont exposés et de pondérer cette exposition en fonction du temps passé dans chacun des espaces.

Implications pour la santé publique et la promotion de la santé

Comme il a été mentionné en introduction, l'étude des effets de milieux sur la santé a deux implications pour la recherche et l'intervention en santé publique et en promotion de la santé : une compréhension de la signification étiologique du milieu pour la santé et une attention dirigée vers les caractéristiques des milieux qui sont favorables ou délétères pour la santé. Les zones homogènes créées pour cette thèse ne doivent pas être perçues comme des nouveaux territoires vers lesquels diriger les actions et politiques de santé publique pour promouvoir la pratique régulière de la marche. Plutôt, la pertinence des zones réside dans l'optimisation de l'examen d'un phénomène de santé de façon à identifier le plus justement possible ses déterminants environnementaux, et ce pour informer les interventions qui visent la modification des milieux pour les rendre plus favorables à la santé.

Toutefois, au niveau de la surveillance, les résultats justifient le besoin de mettre en place des protocoles d'acquisition de données stratifiées et d'inventorier les facteurs écologiques d'exposition au niveau local. Plus particulièrement, il faut spatialiser les données de surveillance car les deux études empiriques démontrent que les expositions dans les unités administratives sont hétérogènes. Cette hétérogénéité pose un défi particulier pour l'intervention : sur un même territoire, une intervention peut être appropriée pour un secteur mais non adéquate pour un autre. Documenter les déterminants environnementaux de la santé, de même que leur changement dans le temps, apparait comme une composante nécessaire des activités de surveillance.

Les résultats de la thèse mettent en relief le bien-fondé de diriger les actions au niveau des environnements (de même qu'au niveau individuel) mais aussi de considérer les multiples échelles au niveau desquelles les facteurs environnementaux influencent la marche. En lien avec la marche comme stratégie d'action pour promouvoir la pratique régulière d'activités physiques dans la population, deux constats peuvent être formulés. Dans un premier temps, l'accessibilité à des services de proximité appert comme un déterminant important de la marche utilitaire. Or, puisque les caractéristiques de l'environnement bâti sont souvent corrélées pour créer différents contextes environnementaux (tel qu'il est illustré par les sept catégories d'exposition), il est peu probable que la pratique de la marche soit accrue uniquement par l'amélioration de l'accessibilité géographique aux services par des modes de transports actifs. D'ailleurs, ce qui émerge des résultats est l'importance de la densité : densité des destinations accessibles par la marche et la densité de population. Mais cette observation est valide uniquement pour la marche totale et la marche utilitaire. En effet, le second constat est lié à la spécificité des associations : les caractéristiques des milieux qui facilitent ou contraignent la pratique de la marche diffèrent selon les motifs de marche. La marche utilitaire est pratiquée plus fréquemment dans les milieux plus urbains alors que la marche récréative semble être pratiquée plus souvent en banlieue. Les interventions à mettre de l'avant pour encourager la marche doivent être adaptées au contexte d'implantation.

À ce titre, l'évaluation des actions visant la modification des environnements doit informer le choix des interventions à implanter pour promouvoir une vie active au niveau populationnel. Une attention plus grande à l'évaluation des impacts sur la santé de la population des politiques publiques et d'expériences naturelles est donc requise. Qui plus est, la formulation et l'implantation de politiques publiques saines pour la santé émergeront vraisemblablement de partenariats et de collaborations intersectorielles entre directions de santé publique, milieux académiques, instances municipales et gouvernementales et les populations locales.

CONCLUSION

Alors que la santé publique connaît une expansion considérable des travaux théoriques et empiriques portant sur le rôle des milieux résidentiels pour la production et le maintien de la santé, l'opérationnalisation de l'espace demeure problématique. Plutôt que de recourir à des unités spatiales administratives de « convenance », l'originalité de cette thèse réside dans l'opérationnalisation du milieu en cohérence avec la distribution spatiale de facteurs d'exposition spécifiques associés avec un indicateur de santé spécifique, de façon à délimiter des unités spatiales homogènes au niveau de l'exposition. Cela s'est opéré à travers l'exemple du potentiel de vie active et les comportements de marche. Les résultats de la thèse confirment empiriquement que le développement et l'application de cette approche ont contribué à la compréhension des effets de milieux sur les comportements de marche, notamment en raison d'une validité interne accrue du devis d'étude et la précision des inférences statistiques qui s'en trouve augmentée. Un schéma décisionnel pour guider l'opérationnalisation des milieux a été proposé.

Comme il a été présenté en introduction, les représentations de l'espace dans la formulation de la problématique de recherche portant sur les effets de milieux sur la santé sont multiples. Chacune de ces conceptualisations répond à des objectifs spécifiques de recherche au regard desquels les milieux doivent être opérationnalisés en cohérence. La conceptualisation de l'espace privilégiée dans cette thèse est celle de l'espace *intégré*, c.-à-d. un milieu dans lequel les individus sont imbriqués et exposés aux mêmes conditions. Toutefois, il est peu probable qu'une seule conceptualisation de l'espace fasse l'unanimité parmi les chercheurs et les décideurs. La compréhension de l'influence du milieu sur la santé est tributaire de la jonction de ses diverses conceptualisations et opérationnalisations qui témoignent de la complexité des mécanismes à travers lesquels le milieu influence la santé.

REFERENCES

Active Living Research http://www.activelivingresearch.org.

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.

Anselin, L., Florax, R. J. G. M., & Rey, S. J. (Eds.) (2004). Advances in Spatial Econometrics: Methodology, Tools and Applications. Berlin Springer-Verlag.

Apparicio, P., Abdelmajid, M., Riva, M., Shearmur, R. (2008). Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics,* 7, 7.

Apparicio, P., Cloutier, M.-S., & Shearmur, R. (2007). The case of Montréal's missing food deserts: Evaluation of accessibility to food supermarkets. *International Journal of Health Geographics,* 6(1), 4.

Apparicio, P., & Séguin, A. M. (2006). Measuring the accessibility of services and facilities for residents of public housing in Montréal. *Urban Studies,* 43(1), 187-211.

Apparicio, P., Shearmur, R., Brochu, M., & Dussault, G. (2003). The measure of distance in a social science policy context: Advantages and costs of using network distances in eight Canadian metropolitan areas. *Journal of Geographic Information and Decision Analysis,* 7(2), 105 - 131.

Arksey, H., & O'Malley, L. (2005). Scoping Studies: towards a methodological framework. *International Journal of Social Research Methodology,* 8(1), 19-32.

Bailey, T. C., & Gatrell, A. C. (1995). Interactive Spatial Data Analysis. Essex: Longman Scientific and Technical.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). Hierarchical modelling and analysis for spatial data. Boca Raton, FA: Chapman & Hall.

Beland, F., Birch, S., & Stoddart, G. (2002). Unemployment and health: contextual-level influences on the production of health in populations. *Social Science & Medicine,* 55(11), 2033-2052.

Berkman, L. F., & Kawachi, I. (Eds.) (2000). Social epidemiology. New York: Oxford University Press.

Bingenheimer, J. B. (2005). Multilevel models and scientific progress in social epidemiology. *Journal of Epidemiology and Community Health,* 59(6), 438-439.

Bingenheimer, J. B., & Raudenbush, S. W. (2004). Statistical and substantive inferences in public health: Issues in the application of multilevel models. *Annual Review of Public Health,* 25, 53-77.

Blakely, T. A., Kennedy, B. P., Glass, R., & Kawachi, I. (2000). What is the lag time between income inequality and health status? *Journal of Epidemiology and Community Health,* 54(4), 318-319.

Blakely, T. A., Kennedy, B. P., & Kawachi, I. (2001). Socioeconomic inequality in voting participation and self-rated health. *American Journal of Public Health,* 91(1), 99-104.

Blakely, T. A., Lochner, K., & Kawachi, I. (2002). Metropolitan area income inequality and self-rated health - a multi-level study. *Social Science & Medicine,* 54(1), 65-77.

Blakely, T. A., & Woodward, A. J. (2000). Ecological effects in multi-level studies. *Journal of Epidemiology and Community Health,* 54(5), 367-374.

Borrell, C., Rodriguez, M., Ferrando, J., Brugal, M. T., Pasarin, M. I., Martinez, V., et al. (2002). Role of individual and contextual effects in injury mortality: new evidence from small area analysis. *Injury Prevention,* 8(4), 297-302.

Bosma, H., van de Mheen, H. D., Borsboom, G., & Mackenbach, J. P. (2001). Neighborhood socioeconomic status and all-cause mortality. *American Journal of Epidemiology,* 153(4), 363-371.

Browning, C. R., & Cagney, K. A. (2002). Neighborhood structural disadvantage, collective efficacy, and self-rated physical health in an urban setting. *Journal of Health and Social Behavior,* 43(4), 383-399.

Browning, C. R., & Cagney, K. A. (2003). Moving beyond poverty: Neighborhood structure, social processes, and health. *Journal of Health and Social Behavior, 44*(4), 552-571.

Browning, C. R., Cagney, K. A., & Wen, M. (2003). Explaining variation in health status across space and time: implications for racial and ethnic disparities in self-rated health. *Social Science & Medicine, 57*(7), 1221-1235.

Cagney, K. A., Browning, C. R., & Wen, M. (2005). Racial disparities in self-rated health at older ages: What difference does the neighborhood make? *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 60*(4), S181-S190.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1-27.

Center for Disease Control and Prevention (2003). Prevalence of physical activity, including lifestyle activities among adults - United States, 2000-2001. *Morbidity and Mortality Weekly Report, 52*(32), 764-769.

Cerin, E., Leslie, E., du Toit, L., Owen, N., & Frank, LD. (2007). Destinations that matter: associations with walking for transport. *Health & Place, 13*(3), 713–724.

Cervero, R. (1996). Mixed land-uses and commuting: Evidence from the American housing survey. *Transportation Research Part a-Policy and Practice, 30*(5), 361-377.

Chaix, B., & Chauvin, P. (2003). Tobacco and alcohol consumption, sedentary lifestyle and overweightness in France: A multilevel analysis of individual and area-level determinants. *European Journal of Epidemiology, 18*(6), 531-538.

Chaix, B., Guilbert, P., & Chauvin, P. (2004). A multilevel analysis of tobacco use and tobacco consumption levels in France - Are there any combination risk groups? *European Journal of Public Health, 14*(2), 186-190.

Chaix, B., Leyland, A. H., Sabel, C. E., Chauvin, P., Rastam, L., Kristersson, H., et al. (2006). Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmö, Sweden, in 2001. *Journal of Epidemiology and Community Health,* 60, 427-435.

Chaix, B., Merlo, J., Subramanian, S. V., Lynch, J., & Chauvin, P. (2005). Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: The case of mental and behavioral disorders due to psychoactive substance use in Malmo, Sweden, 2001. *American Journal of Epidemiology,* 162(2), 171-182.

Chaix, B., Rosvall, M., Lynch, J., & Merlo, J. (2006). Disentangling contextual effects on cause specific mortality in a longitudinal 23-year follow-up study: impact of population density or socioeconomic environment? . *International Journal of Epidemiology,* 35, 633-643.

Chang, V. W., & Christakis, N. A. (2005). Income inequality and weight status in US metropolitan areas. *Social Science & Medicine,* 61(1), 83-96.

Chuang, Y. C., Cubbin, C., Ahn, D., & Winkleby, M. A. (2005). Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. *Journal of Epidemiology and Community Health,* 59(7), 568-573.

Cliff, A., & Ord, J. (1973). Spatial autocorrelation. In R. Chorley & D. Harvey (Eds.), Monograph No. 5 in spatial and environmental systems analysis (p. 178). London: Pion Ltd.

Cloutier, M. S., Apparicio, P., & Thouez, J. P. (2007). GIS-based spatial analysis of child pedestrian accidents near primary schools in Montréal, Canada. *Applied GIS,* 3(4).

Cockings, S., & Martin, D. (2005). Zone design for environment and health studies using pre-aggregated data. *Social Science & Medicine,* 60(12), 2729-2742.

Communauté urbaine de Montréal (2001). Carte d'occupation du sol (édition 2000); document d'accompagnement. Montréal: Communauté urbaine de Montréal, division de l'aménagement, service de la mise en valeur du territoire.

Coulton, C. J., Korbin, J., Chan, T., & Su, M. (2001). Mapping residents' perceptions of neighborhood boundaries: A methodological note. *American Journal of Community Psychology,* 29(2), 371-383.

Craig, C. L., Marshall, A. L., Sjostrom, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., et al. (2003). International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc,* 35, 1381-1395.

Craig, N. (2005). Exploring the generalisability of the association between income inequality and self-assessed health. *Social Science & Medicine,* 60(11), 2477-2488.

Cubbin, C., & Winkleby, M. A. (2005). Protective and harmful effects of neighborhood-level deprivation on individual-level health knowledge, behavior changes, and risk of coronary heart disease. *American Journal of Epidemiology,* 162(6), 559-568.

Cummins, S., Curtis, S., Diez-Roux, A. V., & Macintyre, S. (2007). Understanding and representing 'place' in health research: A relational approach. *Social Science & Medicine,* 65(9), 1825-1838.

Cummins, S., Macintyre, S., Davidson, S., & Ellaway, A. (2005). Measuring neighbourhood social and material context: generation and interpretation of ecological data from routine and non-routine sources. *Health & Place,* 11(3), 249-260.

Cummins, S., Petticrew, M., Higgins, C., Findlay, A., & Sparks, L. (2005). Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. *Journal of Epidemiology and Community Health,* 59(12), 1035-1040.

Cummins, S., Stafford, M., Macintyre, S., Marmot, M., & Ellaway, A. (2005). Neighbourhood environment and its association with self rated health: evidence from Scotland and England. *Journal of Epidemiology and Community Health,* 59(3), 207-213.

Curtis, S., & Jones, I. R. (1998). Is there a place for geography in the analysis of health inequality? *Sociology of Health & Illness,* 20(5), 645-672.

Curtis, S., Southall, H., Congdon, P., & Dodgeon, B. (2004). Area effects on health variation over the life-course: analysis of the longitudinal study sample in area of residence England using new data on in childhood. *Social Science & Medicine,* 58(1), 57-74.

Daniel, M., Moore, S., & Kestens, Y. (2008). Framing the biosocial pathways underlying associations between place and cardiometabolic disease. *Health and Place,* 14(2), 117-132.

De Bourdeaudhuij, I., Sallis, J. F., & Saelens, B. E. (2003). Environmental correlates of physical activity in a sample of Belgian adults. *American Journal of Health Promotion,* 18(1), 83-92.

Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health,* 21, 171-192.

Diez-Roux, A. V. (2003). The examination of neighbourhood effects on health: conceptual and methodological issues related to the presence of multiple levels of organisations. In I. Kawachi & L. F. Berkman (Eds.), Neighborhoods and health pp. 25-64). New York: Oxford University Press.

Diez-Roux, A. V. (2004a). Estimating neighborhood health effects: the challenges of causal inference in a complex world - Commentary. *Social Science & Medicine,* 58(10), 1953-1960.

Diez-Roux, A. V. (2004b). The study of group-level factors in epidemiology: Rethinking variables, study designs, and analytical approaches. *Epidemiologic Reviews,* 26, 104-111.

Diez-Roux, A. V., Link, B. G., & Northridge, M. E. (2000). A multilevel analysis of income inequality and cardiovascular disease risk factors. *Social Science & Medicine,* 50(5), 673-687.

Diez-Roux, A. V., Nieto, F. J., Caulfield, L., Tyroler, H. A., Watson, R. L., & Szklo, M. (1999). Neighbourhood differences in diet: the Atherosclerosis Risk in Communities (ARIC) Study. *Journal of Epidemiology and Community Health,* 53(1), 55-63.

Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health,* 91(11), 1783-1789.

DMTI Spatial (2005). CanMap® Streetfiles. Markham: DMTI Spatial inc.

Drukker, M., Kaplan, C., Feron, F., & van Os, J. (2003). Children's health-related quality of life, neighbourhood socio-economic deprivation and social capital. A contextual analysis. *Social Science & Medicine,* 57(5), 825-841.

Drukker, M., Kaplan, C., & van Os, J. (2005). Residential instability in socioeconomically deprived neighbourhoods, good or bad? *Health & Place,* 11(2), 121-129.

Drukker, M., & van Os, J. (2003). Mediators of neighbourhood socioeconomic deprivation and quality of life. *Social Psychiatry and Psychiatric Epidemiology,* 38(12), 698-706.

Duda, R., Hart, P., & Stork, D. (2001). Pattern classification. New York: John Wiley & Sons, Ltd.

Duncan, C., Jones, K., & Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. *Social Science & Medicine,* 46(1), 97-117.

Duncan, C., Jones, K., & Moon, G. (1999). Smoking and deprivation: are there neighbourhood effects? *Social Science & Medicine,* 48(4), 497-505.

Dunn, A. L., Marcus, B. H., Kampert, J. B., Garcia, M. E., Kohl, H. W., & Blair, S. N. (1999). Comparison of lifestyle and structured interventions to increase physical activity and cardiorespiratory fitness: A randomized trial. *Journal of the American Medical Association,* 281, 327-334.

Ecob, R., & Macintyre, S. (2000). Small area variations in health related behaviours; do these depend on the behaviour itself, its measurement, or on personal characteristics? *Health & Place,* 6(4), 261-274.

Ellen, I. G., Mijanovich, T., & Dillman, K.-N. (2001). Neighborhood Effects on Health: Exploring the Links and Assessing the Evidence. *Journal of Urban Affairs,* 23(3&4), 391-408.

ESRI. (2006). ArcGIS 9.2. Redlands, CA.

Everitt, B., Landau, S., & Leese, M. (2001). Cluster analysis. London: Arnold.

Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., & Raudenbush, S. (2003). Relationship between urban sprawl and physical activity, obesity, and morbidity. *American Journal of Health Promotion,* 18(1), 47-57.

Eyler, A. A., Brownson, R. C., Bacak, S. J., & Housemann, R. A. (2003). The epidemiology of walking for physical activity in the United States. *Medicine and Science in Sports and Exercise,* 35(9), 1529-1536.

Fisher, K., Li, F., & Michael, Y. (2004). Neighborhood-level influences on physical activity among older adults: a multi-level analysis. *Journal of Aging and Physical Activity,* 12(1), 45-63.

Flowerdew, R., Manley, D., & Sabel, C. (2008) Neighbourhood effects on health: Does it matter where you draw the boundaries? *Social Science & Medicine,* 66(6), 1241-1255.

Forsyth, A., Oakes, J. M., Schmitz, K. H., & Hearst, M. (2007). Does residential density increase walking and other physical activity? *Urban Studies,* 44(4), 679-697.

Fortin, M.-J., & Dale, M. (2005). Spatial Analysis: A Guide for Ecologists. Cambridge: Cambridge University Press.

Fotheringham, A. S., Brundson, C., & Charlton, M. (1996). he geography of parameter space: an investigation into spatial nonstationnary. *International Journal of Geographical Information Systems,* 10(5), 605-628.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). Quantitative geography. Perspectives on spatial data analysis. London: Sage.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). Geographically weighted regression: The analysis of spatially varying relationships. Chichester, UK: John Wiley & Sons Ltd.

Frank, L. D., & Pivo, G. (1995). Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking. *Transportation Research Record,* 1466, 44-52.

Frank, L. D., Saelens, B. E., Powell, K. E., & Chapman, J. E. (2007). Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Social Science & Medicine,* 65 1898-1914.

Frank, L. D., Schmid, T. L., Sallis, J. F., Chapman, J., & Saelens, B. E. (2005). Linking objectively measured physical activity with objectively measured urban form - Findings from SMARTRAQ. *American Journal of Preventive Medicine,* 28(2), 117-125.

Franzini, L., Caughy, M., Spears, W., & Esquer, M. E. F. (2005). Neighborhood economic conditions, social processes, and selfrated health in low-income neighborhoods in Texas: A multilevel latent variables model. *Social Science & Medicine,* 61(6), 1135-1150.

Franzini, L., & Spears, W. (2003). Contributions of social context to inequalities in years of life lost to heart disease in Texas, USA. *Social Science & Medicine,* 57(10), 1847-1861.

Frohlich, K. L., Potvin, L., Chabot, P., & Corin, E. (2002). A theoretical and empirical analysis of context: neighbourhoods, smoking and youth. *Social Science & Medicine,* 54, 1401-1417.

Frohlich, K. L., Potvin, L., Gauvin, L., & Chabot, P. (2002). Youth smoking initiation: disentangling context from composition. *Health & Place,* 8(3), 155-166.

Galea, S., & Ahern, J. (2006). Invited commentary: Considerations about specificity of associations, causal pathways, and heterogeneity in multilevel thinking. *American Journal of Epidemiology,* 163(12), 1079-1082.

Galster, G. (2001). On the nature of neighbourhood. *Urban Studies,* 38(12), 2111-2124.

Gauvin, L., Richard, L., Craig, C. L., Spivock, M., Riva, M., Forster, M., et al. (2005). From walkability to active living potential - An "ecometric" validation study. *American Journal of Preventive Medicine,* 28(2), 126-133.

Gauvin, L., Riva, M., Barnett, T. A., Richard, L., Craig, C. L., Spivock, M., et al. Association between Neighborhood Active Living Potential and Walking. *American Journal of Epidemiology,* 167(8), 944-953.

Gauvin, L., Robitaille, E., Riva, M., McLaren, L., Dassa, C., & Potvin, L. (2007). Conceptualizing and operationalizing neighbourhoods: the conundrum of identifying territorial units. *Canadian Journal of Public Health,* 98(suppl 1), S18-S26.

Gee, G. C. (2002). A multilevel analysis of the relationship between institutional and individual racial discrimination and health status. *American Journal of Public Health,* 92(4), 615-623.

Gee, G. C., & Takeuchi, D. T. (2004). Traffic stress, vehicular burden and well-being: A multilevel analysis. *Social Science & Medicine,* 59(2), 405-414.

Giles-Corti, B., Broomhall, M. H., Knuiman, M., & al., e. (2005). Increasing walking: how important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine,* 28(Suppl 2), 169-176.

Giles-Corti, B., & Donovan, R. J. (2002a). The relative influence of individual, social and physical environment determinants of physical activity. *Social Science & Medicine,* 54(12), 1793-1812.

Giles-Corti, B., & Donovan, R. J. (2002b). Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. *Preventive Medicine,* 35(6), 601-611.

Giles-Corti, B., & Donovan, R. J. (2003). Relative influences of individual, social environmental, and physical environmental correlates of walking. *American Journal of Public Health,* 93(9), 1583-1589.

Hakim, A. A., Petrovitch, H., Burchfiel, C. M., Ross, G. W., Rodriguez, B. L., White, L. R., et al. (1998). Effects of walking on mortality among nonsmoking retired men. *New England Journal of Medicine,* 338, 94-99.

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the Built Environment Affects Physical Activity. Views from Urban Planning. *American Journal of Preventive Medicine,* 23(2S), 64 -73.

Handy, S. L., Cao, X., & Mokhtarian, P. L. (2006). Does self-selection explain the relationship between built environment and walking behavior? Empirical evidence from Northern California. *Journal of the American planning association,* 72(1), 55-74.

Harris, R., Sleight, P., & Webber, R. (2005). Geodemographics, GIS and neighbourhood targetting. Chichester, West Sussex, England: John Wiley & Sons, Ltd.

Harris, R. J., & Longley, P. A. (2002). Creating small area measures of urban deprivation. *Environment and Planning A,* 34(6), 1073-1093.

Haynes, R., Jones, A., Reading, R., Daras, K., & Emond, A. (2008). Neighbourhood variations in child accidents and related child and maternal characteristics: Does area definition make a difference? *Health & Place,* 14(4), 693-701.

Haynes, R., Daras, K., Reading, R., & Jones, A. (2007). Modifiable neighbourhood units, zone design and residents ' perceptions. *Health & Place,* 13(4), 812-825.

Haynes, R., & Gale, S. (1999). Mortality, long-term illness and deprivation in rural and metropolitan wards of England and Wales. *Health & Place,* 5(4), 301-312.

Hembree, C., Galea, S., Ahern, J., Tracy, M., Piper, T. M., Miller, J., et al. (2005). The urbant built environment and overdose mortality in New York City neighborhoods. *Health & Place,* 11(2), 147-156.

Hewko, J., Smoyer-Tomic, K. E., & Hodgson, M. J. (2002). Measuring neighbourhood spatial accessibility to urban amenities: Does aggregation error matter? *Environment and Planning A,* 34(7), 1185 - 1206.

Hopman, W. M., Berger, C., Joseph, L., Towheed, T., Anastassiades, T., Tenenhouse, A., et al. (2002). Is there regional variation in the SF-36 scores of Canadian adults? *Canadian Journal of Public Health-Revue Canadienne De Sante Publique,* 93(3), 233-237.

Hou, F., & Chen, J. (2003). Neighbourhood low income, income inequality and health in Toronto. *Health Reports,* 14, 21-34.

Hou, F., & Myles, J. (2005). Neighbourhood inequality, neighbourhood affluence and population health. *Social Science & Medicine,* 60(7), 1557-1569.

Hu, F., Siga, R., Rich-Edwards, J., Colditz, G., Solomon, C., Willet, W., et al. (1999). Walking compared with vigorous physical activity and risk of type-2 diabetes in women. *Journal of the American Medical Association,* 282, 1433-1439.

Humpel, N., Owen, N., Leslie, E., Marshall, A. L., Bauman, A. E., & Sallis, J. F. (2004). Associations of location and perceived environmental attributes with walking in neighborhoods. *American Journal of Health Promotion,* 18(3), 239-242.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior,* 38(1), 21-37.

Jaffe, D. H., Eisenbach, Z., Neumark, Y. D., & Manor, O. (2005a). Does living in a religiously affiliated neighborhood lower mortality? *Annals of Epidemiology,* 15(10), 804-810.

Jaffe, D. H., Eisenbach, Z., Neumark, Y. D., & Manor, O. (2005b). Individual, household and neighborhood socioeconomic status and mortality: a study of absolute and relative deprivation. *Social Science & Medicine,* 60(5), 989-997.

Jerrett, M., Burnet, R. T., Ma, R., Pope, C. A., Krewski, D., Newbold, B., et al. (2006). Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology,* 17(6), S69-S69.

Jones, K., & Moon, G. (1993). Medical geography: taking space seriously. *Progress in Human Geography,* 17(4), 515-524.

Kairouz, S., Nadeau, L., & Siou, G. L. (2005). Area variations in the prevalence of substance use and gambling behaviours and problems in Québec: a multilevel analysis. *Canadian Journal of Psychiatry,* 50(10), 591-598.

Kaplan, G. A. (2004). What's wrong with social epidemiology, and how can we make it better? *Epidemiologic Reviews,* 26, 124-135.

Karlsen, S., Nazroo, J. Y., & Stephenson, R. (2002). Ethnicity, environment and health: putting ethnic inequalities in health in their place. *Social Science & Medicine,* 55(9), 1647-1661.

Kavanagh, A. M., Goller, J. L., King, T., Jolley, D., Crawford, D., & Turrell, G. (2005). Urban area disadvantage and physical activity: a multilevel study in Melbourne, Australia. *Journal of Epidemiology and Community Health,* 59(11), 934-940.

Kearns, R. A. (1993). Place and health: Towards a reformed medical geography. *Professional Geographer,* 45, 139-147.

Kearns, R. A., & Joseph, A. E. (1993). Space in its place: Developing the link in medical geography. *Social Science and Medicine,* 37, 711-717.

Kemp, K. K. (Ed.) (2008). Encyclopedia of geographic information science. Thousand Oaks, CA: Sage.

Kerr, J., Rosenberg, D., Sallis, JF., Saelens, BE., Frank, LD, & Conway, TL. (2006). Active commuting to school: Associations with environment and parental concerns. *Medicine in Science Sports & Exercise,* 38(4), 787-794.

King, W. C., Brach, J. S., Belle, S., Killingsworth, R., Fenton, M., & Kriska, A. M. (2003). The relationship between convenience of destinations and walking levels in older women. *American Journal of Health Promotion,* 18, 74-82.

Krieger, N. (2001). Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology,* 30(4), 668-677.

Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M. J., Subramanian, S. V., & Carson, R. (2002). Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. *American Journal of Epidemiology,* 156(5), 471-482.

Krizek, K. J., & Johnson, P. J. (2006). Proximity to trails and retail: Effects on urban cycling and walking. *Journal of the American planning association,* 72(1), 33-42.

Law, M., Wilson, K., Eyles, J., Elliott, S., Jerrett, M., Moffat, T., et al. (2005). Meeting health need, accessing health care: the role of neighbourhood. *Health & Place,* 11(4), 367-377.

Lawson, A. B., & Denison, D. G. T. (Eds.) (2002). Spatial Cluster Modelling. Boca Raton, FA: Chapman & Hall/CRC.

Lee, C., & Vernez Moudon, A. (2006). Correlates of Walking for Transportation or Recreation Purposes. *Journal of Physical Activity and health,* 3 (Suppl 1), S77.

Lee, I., & Paffenbarger, R. J. (2000). Association of light, moderate, and vigorous intensity physical activity with longevity: The Harvard Alumni Study. *American Journal of Epidemiology,* 151, 293-299.

Lee, J., & Wong, D. W. S. (2001). Statistical analysis with ArcView GIS. New York: John Wiley & Sons, Inc.

Leslie, E., Coffee, N., Frank, L. D., Owen, N., Bauman, A., & Hugo, G. (2007). Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. *Health & Place,* 13(1), 111–122.

Leventhal, T., & Brooks-Gunn, J. (2003). Moving to opportunity: an experimental study of neighborhood effects on mental health. *American Journal of Public Health,* 93(9), 1576-1582.

Leyland, A. H. (2005). Socioeconomic gradients in the prevalence of cardiovascular disease in Scotland: the roles of composition and context. *Journal of Epidemiology and Community Health,* 59(9), 799-803.

Li, F. Z., Fisher, K. J., Bauman, A., Ory, M. G., Chodzko-Zajko, W., Harmer, P., et al. (2005). Neighborhood influences on physical activity in middle-aged and older adults: A multilevel perspective. *Journal of Aging and Physical Activity,* 13(1), 87-114.

Li, F. Z., Fisher, K. J., Brownson, R. C., & Bosworth, M. (2005). Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *Journal of Epidemiology and Community Health,* 59(7), 558-564.

Lindstrom, M., Moghaddassi, M., Bolin, K., Lindgren, B., & Merlo, J. (2003). Social participation, social capital and daily tobacco smoking: a population-based multilevel analysis in Malmo, Sweden. *Scandinavian Journal of Public Health,* 31(6), 444-450.

Lindstrom, M., Moghaddassi, M., & Merlo, J. (2003). Social capital and leisure time physical activity: a population based multilevel analysis in Malmo, Sweden. *Journal of Epidemiology and Community Health,* 57(1), 23-28.

Lindstrom, M., Moghaddassi, M., & Merlo, J. (2004). Individual self-reported health, social participation and neighbourhood: a multilevel analysis in Malmo, Sweden. *Preventive Medicine,* 39(1), 135-141.

Lochner, K. A., Kawachi, I., Brennan, R. T., & Buka, S. L. (2003). Social capital and neighborhood mortality rates in Chicago. *Social Science & Medicine,* 56(8), 1797-1805.

Longley, P., & Batty, M. (1996). Spatial analysis: Modelling in a GIS environment. Cambridge: earson Professional Ltd.

Lovasi, G. S., Vernez-Moudon, A., Pearson, A. L., Hurvitz, P. M., Larson, E. B., Siscovick, D. S., et al. (2008). Using built environment characteristics to predict walking for exercise. *International Journal of Health Geographics,* 7:10.

Macintyre, S., & Ellaway, A. (2000). Ecological approaches: rediscovering the role of the physical and social environment. In L. F. Berkman & I. Kawachi (Eds.), Social epidemiology pp. 332-348). New York: Oxford University Press.

Macintyre, S., & Ellaway, A. (2003). Neighborhoods and health: an overview. In I. Kawachi & L. F. Berkman (Eds.), Neighborhoods and Health pp. 20-42). New York: Oxford University Press.

Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science & Medicine,* 55(1), 125-139.

Macintyre, S., McIver, S., & Sooman, A. (1993). Area, class and health: Should we be focusing on people or places? *Journal of Social Policy,* 22, 213-234.

Malmstrom, M., Sundquist, J., & Johansson, S. E. (1999). Neighborhood environment and self-reported health status: A multilevel analysis. *American Journal of Public Health,* 89(8), 1181-1186.

Marinacci, C., Spadea, T., Biggeri, A., Demaria, M., Caiazzo, A., & Costa, G. (2004). The role of individual and contextual socioeconomic circumstances on mortality: analysis of time variations in a city of north west Italy. *Journal of Epidemiology and Community Health,* 58(3), 199-207.

Martikainen, P., Kauppinen, T. M., & Valkonen, T. (2003). Effects of the characteristics of neighbourhoods and the characteristics of people on cause specific mortality: a register based follow up study of 252,000 men. *Journal of Epidemiology and Community Health,* 57(3), 210-217.

Martin, D., Nolan, A., & Tranmer, M. (2001). The application of zone-design methodology in the 2001 UK Census. *Environment and Planning A,* 33(11), 1949-1962.

Mays, N., Roberts, E., & Popay, J. (2001). Synthesising research evidence. In A. P. Fulop N, Clarke A, Black N (Ed.), Studying the organisation and delivery of health services: research methods. London: Routledge

McCulloch, A. (2001). Ward-level deprivation and individual social and economic outcomes in the British Household Panel Study. *Environment and Planning A,* 33(4), 667-684.

Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., et al. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health,* 60(4), 290-297.

Merlo, J., Chaix, B., Yang, M., Lynch, J., & Rastam, L. (2005). A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiology and Community Health,* 59(12), 1022-1028.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika,* 50(2), 159-179.

Mohan, J., Twigg, L., Barnard, S., & Jones, K. (2005). Social capital, geography and health: a small-area analysis for England. *Social Science & Medicine,* 60(6), 1267-1283.

Morland, K., Wing, S., & Roux, A. D. (2002). The contextual effect of the local food environment on residents' diets: The atherosclerosis risk in communities study. *American Journal of Public Health,* 92(11), 1761-1767.

Morris, R. W., Whincup, P. H., Lampe, F. C., Walker, N., Wannamethee, S. G., & Shaper, A. G. (2001). Geographic variation in incidence of coronary heart disease in Britain: the contribution of established risk factors. *Heart,* 86(3), 277-283.

Murray, A. T. (1999). Spatial analysis using clustering methods: evaluating central point and median approaches. *Journal of Geographical Systems,* 1(4), 367-383.

Murray, A. T., & Estivill-Castro, V. (1998). Cluster discovery techniques for exploratory spatial data analysis. *International Journal of Geographical Information Science,* 12(5), 431-443.

Nakaya, T. (2000). An information statistical approach to the modifiable areal unit problem in incidence rate maps. *Environment and Planning A,* 32(1), 91-109.

Nakaya, T., Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine,* 24(17), 2695-2717.

Nolin, B., & Hamel, D. (2005). Plan d'action Kino-Québec 2005-2008, chapitre 2, texte de base (document non publié). Institut national de santé publique du Québec.

O'Campo, P. (2003). Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health. *American Journal of Epidemiology,* 157(1), 9-13.

Oakes, J. M. (2004a). Causal inference and the relevance of social epidemiology - Rejoinder. *Social Science & Medicine,* 58(10), 1969-1971.

Oakes, J. M. (2004b). The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social Science & Medicine,* 58(10), 1929-1952.

Oakes, J. M., & Johnson, P. H. (2006). Propensity score matching for social epidemiology. In J. M. Oakes & J. S. Kaufman (Eds.), Methods in social epidemiology pp. 364-386). San Francisco: Jossey-Bass.

Oakes, J. M., & Rossi, P. H. (2003). The measurement of SES in health research: current practice and steps toward a new approach. . *Social Science & Medicine,* 56(4), 769-784.

Oliver, L., & Hayes, M. V. (2007). Does choice of spatial unit matter for estimating small area disparities in health and place effects in the Vancouver census metropolitain area? *Canadian Journal of Public Health,* 98, S27-S34.

Openshaw, S. (1977). A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the Institute of British Geographers,* 2, 459-472.

Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography,* 38, 1-41.

Openshaw, S. (1996). Developing GIS-relevant zone-based spatial analysis methods. In P. Longley & M. Batty (Eds.), Spatial analysis: Modelling in a GIS environment. pp. 55-73). Cambridge: Pearson Professional Ltd.

Openshaw, S., & Rao, L. (1995). Algorithms for reengineering 1991 census geography. *Environment and Planning A,* 27(3), 425-446.

Openshaw, S., & Taylor, P. J. (1979). A Million or So Correlated Coefficients: Three experiments on the Modifiable Areal Unit Problem. In N. Wrigley & R. Bennet (Eds.), Statistical applications in the spatial sciences pp. 127-144). London: Pion.

Osypuk, T. L., & Galea, S. (2007). What level macro? Choosing appropriate level to assess how place influences population health. In S. Galea (Ed.), Macrosocial determinants of population health pp. 399-438). New York: Springer.

Owen, N., Humpel, N., Leslie, E., Bauman, A., & Sallis, J. F. (2004). Understanding environmental influences on walking - Review and research agenda. *American Journal of Preventive Medicine,* 27(1), 67-76.

Pampalon, R., Duncan, C., Subramanian, S. V., & Jones, K. (1999). Geographies of health perception in Quebec: a multilevel perspective. *Social Science & Medicine,* 48(10), 1483-1490.

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., et al. (1995). Physical activity and public health - a recommendation from the Centers for disease control and prevention and the American college of sports medicine. *Jama-Journal of the American Medical Association,* 273(5), 402-407.

Patel, K. V., Eschbach, K., Rudkin, L. L., Peek, M. K., & Markides, K. S. (2003). Neighborhood context and self-rated health in older Mexican Americans. *Annals of Epidemiology,* 13(9), 620-628.

Pearce, J., Witten, K., & Bartie, P. (2006). Neighbourhoods and health: A GIS approach to measuring community resource accessibility. *Journal of Epidemiology and Community Health,* 60(5), 389-395.

Petticrew, M., Cummins, S., Ferrell, C., Findlay, A., Higgins, C., Hoy, C., et al. (2005). Natural experiments: an underused tool for public health. *Public Health,* 119, 751-757.

Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology and Community Health,* 55(2), 111-122.

Pikora, T. J., Giles-Corti, B., Knuiman, M. W., Bull, F. C., Jamrozik, K., & Donovan, R. J. (2006). Neighborhood environmental factors correlated with walking near home: Using SPACES. *Medicine and Science in Sports and Exercise,* 38(4), 708-714.

Popay, J., Thomas, C., Williams, G., Bennett, S., Gattrell, A., & Bostock, L. (2003). A proper place to live: Health inequalities, agency and the normative dimensions of space. *Social Science & Medicine,* 57(1), 55-69.

Propper, C., Jones, K., Bolster, A., Burgess, S., Johnston, R., & Sarker, R. (2005). Local neighbourhood and mental health: Evidence from the UK. *Social Science & Medicine,* 61(10), 2065-2083.

Rafferty, A. P., Reeves, M. J., McGee, H. B., & Pivarnik, J. M. (2002). Physical activity patterns among walkers and compliance with public health recommendations. *Medicine and Science in Sports and Exercise,* 34(8), 1255-1261.

Raudenbush, S. W., & Bryk, A. (2002). Hierarchical linear models: applications and data analysis methods Newbury Park, CA: Sage.

Raudenbush, S. W., Bryk, A., & Congdon, R. (2005). HLM: Hierarchical Linear and Nonlinear Modeling (Version 6.04) Chicago: Scientific Software International.

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology,* 29, 1-41.

Reading, R., Langford, I. H., Haynes, R., & Lovett, A. (1999). Accidents to preschool children: comparing family and neighbourhood risk factors. *Social Science & Medicine,* 48(3), 321-330.

Reijneveld, S. A. (2002). Neighbourhood socioeconomic context and self reported health and smoking: a secondary analysis of data on seven cities. *Journal of Epidemiology and Community Health,* 56(12), 935-942.

Reijneveld, S. A., Verheij, R. A., & de Bakker, D. H. (2000). The impact of area deprivation on differences in health: does the choice of the geographical classification matter? *Journal of Epidemiology and Community Health,* 54(4), 306-313.

Riva, M., Gauvin, L., & Barnett, T. A. (2007). Toward the next generation of research into small area effects on health: A synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology and Community Health,* 61(10), 853-861.

Robert, S. A., & Lee, K. Y. (2002). Explaining race differences in health among older adults - The contribution of community socioeconomic context. *Research on Aging,* 24(6), 654-683.

Robert, S. A., & Reither, E. N. (2004). A multilevel analysis of race, community disadvantage, and body mass index among adults in the US. *Social Science & Medicine,* 59(12), 2421-2434.

Roos, L. L., Magoon, J., Gupta, S., Chateau, D., & Veugelers, P. J. (2004). Socioeconomic determinants of mortality in two Canadian provinces: Multilevel modelling and neighborhood context. *Social Science & Medicine,* 59(7), 1435-1447.

Rosen, G. (1993). A history of public health. Baltimore: Johns Hopkins University Press.

Ross, C. (2000). Walking, exercising, and smoking: Does neighbourhood matter? . *Social Science and Medicine,* 51, 265-274.

Ross, N. A., Tremblay, S., & Graham, K. (2004). Neighbourhood influences on health in Montreal, Canada. *Social Science & Medicine,* 59(7), 1485-1494.

Rothman, K., & Greenland, S. (1998). Modern epidemiology. Philadelphia: Lippincott Williams & Wilkins.

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine,* 25(2), 80-91.

Sallis, J. E., Cervero, R. B., Ascher, W., Henderson, K. A., Kraft, M. K., & Kerr, J. (2006). An ecological approach to creating active living communities. *Annual Review of Public Health,* 27, 297-322.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology,* 28, 443-478.

SAS Institute Inc. SAS Technical Report A-108, Cubic clustering criterion. . (p. 54). Cary (NC), USA: SAS Institute Inc.

SAS Institute Inc. SAS version 9.1. Cary, NC, USA.

Satariano, W., & McAuley, E. (2003). Promoting physical activity among older adults: from ecology to the individual. *Am J Prev Med,* 25(3 Suppl 2), 184-192.

Scribner, R. A., Cohen, D. A., & Fisher, W. (2000). Evidence of a structural effect for alcohol outlet density: A multilevel analysis. *Alcoholism-Clinical and Experimental Research,* 24(2), 188-195.

Siegel, P. Z., Brackbill, R. M., & Heath, G. W. (1995). The epidemiology of walking for exercise: implications for promoting activity among sedentary groups. *American Journal of Public Health,* 85, 706-710.

Smoyer-Tomic, K. E., Hewko, J. N., & Hodgson, M. J. (2004). Spatial accessibility and equity of playgrounds in Edmonton, Canada. *The Canadian Geographer,* 48(3), 287-302.

Snijders, T. A. B., & Bosker, R. J. (1999). Multilevel analysis: an introduction to basic and advanced multilevel modeling. Thousand Oaks, CA: Sage.

Stafford, M., Bartley, M., Mitchell, R., & Marmot, M. (2001). Characteristics of individuals and characteristics of areas: investigating their influence on health in the Whitehall II study. *Health & Place,* 7(2), 117-129.

Stafford, M., Cummins, S., Macintyre, S., Ellaway, A., & Marmot, M. (2005). Gender differences in the associations between health and neighbourhood environment. *Social Science & Medicine,* 60(8), 1681-1692.

Stafford, M., & Marmot, M. (2003). Neighbourhood deprivation and health: does it affect us all equally? *International Journal of Epidemiology,* 32(3), 357-366.

Stafford, M., Martikainen, P., Lahelma, E., & Marmot, M. (2004). Neighbourhoods and self rated health: a comparison of public sector employees in London and Helsinki. *Journal of Epidemiology and Community Health,* 58(9), 772-778.

Stampfer, M., Hu, F., Manson, J., Rimm, E., & Willet, W. (2000). Primary prevention of coronary heart disease in women through diet and exercise. *New England Journal of Medicine,* 343, 16-32.

Statistics Canada. 2001 Census preview of products and Services. Ottawa: Statistics Canada.

Statistics Canada Census Operations Division. (2003). 2001 Census Dictionary. Ottawa, Canada: Minister of Industry.

Stjarne, M. K., de Leon, A. P., & Hallqvist, J. (2004). Contextual effects of social fragmentation and material deprivation on risk of myocardial infarction - results from the Stockholm Heart Epidemiology Program (SHEEP). *International Journal of Epidemiology,* 33(4), 732-741.

Subramanian, S. V. (2004). The relevance of multilevel statistical methods for identifying causal neighborhood effects - Commentary. *Social Science & Medicine,* 58(10), 1961-1967.

Subramanian, S. V., Acevedo-Garcia, D., & Osypuk, T. L. (2005). Racial residential segregation and geographic heterogeneity in black/white disparity in poor self-rated health in the US: a multilevel statistical analysis. *Social Science & Medicine,* 60(8), 1667-1679.

Subramanian, S. V., Chen, J. T., Rehkopf, D. H., Waterman, P. D., & Krieger, N. (2005). Racial disparities-in context: A multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. *American Journal of Public Health,* 95(2), 260-265.

Subramanian, S. V., Jones, K., & Duncan, C. (2003). Multilevel methods for public health research. In I. Kawachi & L. F. Berkman (Eds.), Neighborhoods and health pp. 65-111). New York: Oxford University Press.

Subramanian, S. V., & Kawachi, I. (2003). The association between state income inequality and worse health is not confounded by race. *International Journal of Epidemiology,* 32(6), 1022-1028.

Subramanian, S. V., Kawachi, I., & Kennedy, B. P. (2001). Does the state you live in make a difference? Multilevel analysis of self-rated health in the US. *Social Science & Medicine,* 53(1), 9-19.

Subramanian, S. V., Kim, D. J., & Kawachi, I. (2002). Social trust and self-rated health in US communities: A multilevel analysis. *Journal of Urban Health-Bulletin of the New York Academy of Medicine,* 79(4), S21-S34.

Subramanian, S. V., Kim, D. J., & Kawachi, I. (2005). Covariation in the socioeconomic determinants of self rated health and happiness: a multivariate multilevel analysis of individuals and communities in the USA. *Journal of Epidemiology and Community Health,* 59(8), 664-669.

Sundquist, J., Malmstrom, M., & Johansson, S. E. (1999). Cardiovascular risk factors and the neighbourhood environment: a multilevel analysis. *International Journal of Epidemiology,* 28(5), 841-845.

Sundquist, K., Malmstrom, M., & Johansson, S. E. (2004). Neighbourhood deprivation and incidence of coronary heart disease: a multilevel study of 2.6 million women and men in Sweden. *Journal of Epidemiology and Community Health,* 58(1), 71-77.

Sundquist, K., Winkleby, M., Ahlen, H., & Johansson, S. E. (2004). Neighborhood socioeconomic environment and incidence of coronary heart disease: A follow-up study of 25,319 women and men in Sweden. *American Journal of Epidemiology,* 159(7), 655-662.

Susser, M., & Susser, E. (1996a). Choosing a future for epidemiology .1. Eras and paradigms. *American Journal of Public Health,* 86(5), 668-673.

Susser, M., & Susser, E. (1996b). Choosing a future for epidemiology .2. From black box to Chinese boxes and eco-epidemiology. *American Journal of Public Health,* 86(5), 674-677.

Tabachnick, B., & Fidel, L. (2001). Using multivariate statistics Needham Heights, MA: Allyn & Bacon.

Takahashi, L. M., Wiebe, D., & Rodriguez, R. (2001). Navigating the time-space context ofHIV and AIDS: daily routines and access to care. *Social Science & Medicine,* 53(7), 845-863.

Tatalovich, Z., Wilson, J. P., Milam, J. E., Jerrett, M., & McConnell, R. (2006). Competing definitions of contextual environments. *International Journal of Health Geographics,* 5, 55.

The Project on Human Development in Chicago Neighborhoods http://www.icpsr. umich.edu/Phdcn.

Theil, H. (1972). Statistical decomposition analysis. Amsterdam: North-Holland.

Theil, H., & Finezza, A. (1971). A note on the measurement of racial integration of schools by means of informational concepts. *Journal of Mathematical Sociology,* 1, 187-194.

Tilt, J. H., Unfried, T. M., & Roka, B. (2007). Using objective and subjective measures of neighborhood greenness and accessible destinations for understanding walking trips and BMI in Seattle, Washington. *American Journal of Health Promotion,* 21(4), Suppl. 371-379.

Tonne, C., Schwartz, J., Mittleman, M., Melly, S., Suh, H., & Goldberg, R. (2005). Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods. *Circulation,* 111(23), 3063-3070.

Transportation Research Board and Institute of Medicine of the National Academies [TRB & IOM]. (2005). Does the built environment influence physical activity? Examining the evidence. Special Report 282. Transportation Research Board. Washington, DC.

TRB & IOM. (2005). Does the built environment influence physical activity? Examining the evidence. Special Report 282. Transportation Research Board. Washington, DC.

Twigg, L., & Moon, G. (2002). Predicting small area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data. *Social Science & Medicine,* 54(6), 931-937.

United States Department of Health and Human Services [USDHHS]. (1996). Physical Activity and health: A report from the Surgeon General. Atlanta, GA: U.S: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

USDHHS. (1996). Physical Activity and health: A report from the Surgeon General. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

van Lenthe, F. J., Brug, J., & Mackenbach, J. P. (2005). Neighbourhood inequalities in physical inactivity: the role of neighbourhood attractiveness, proximity to local facilities and safety in the Netherlands. *Social Science & Medicine,* 60(4), 763-775.

van Lenthe, F. J., & Mackenbach, J. P. (2002). Neighbourhood deprivation and overweight: the GLOBE study. *International Journal of Obesity,* 26(2), 234-240.

Veenstra, G. (2005). Location, location, location: contextual and compositional health effects of social capital in British Columbia, Canada. *Social Science & Medicine,* 60(9), 2059-2071.

Veugelers, P. J., Yip, A. M., & Kephart, G. (2001). Proximate and contextual socioeconomic determinants of mortality: Multilevel approaches in a setting with universal health care coverage. *American Journal of Epidemiology,* 154(8), 725-732.

Waitzman, N. J., Smith, K. R., & Stroup, A. (1999). The direct and indirect effects of metropolitan area inequality on mortality - A hierarchical analysis. Socioeconomic Status and Health in Industrial Nations pp. 347-349).

WCRF / AICR. (2007). Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR.

Wen, M., Browning, C. R., & Cagney, K. A. (2003). Poverty, affluence, and income inequality: neighborhood economic structure and its implications for health. *Social Science & Medicine,* 57(5), 843-860.

Wendel-Vos, G. C. W., Schuit, A. J., De Niet, R., Boshuizen, H. C., Saris, W. H. M., & Kromhout, D. (2004). Factors of the physical environment associated with walking and bicycling. *Medicine and Science in Sports and Exercise,* 36(4), 725-730.

Wiles, J. (2003). Daily geographies of caregivers: mobility, routine, scale. *Social Science & Medicine,* 57(7), 1307-1325.

Witten, K., Exeter, D., & Field, A. (2003). The quality of urban environments: Mapping variation in access to community resources. *Urban Studies,* 40(1), 161-177.

World Cancer Research Fund / American Institute for Cancer Research (WCRF / AICR). (2007). Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR.

Xi, G. L., McDowell, I., Nair, R., & Spasoff, R. (2005). Income inequality and health in Ontario - A multilevel analysis. *Canadian Journal of Public Health-Revue Canadienne De Sante Publique,* 96(3), 206-211.

Yen, I. H., & Kaplan, G. A. (1999). Neighborhood social environment and risk of death: Multilevel evidence from the Alameda County Study. *American Journal of Epidemiology,* 149(10), 898-907.

ANNEXES

Annexe I

Article 1 : Tableaux 2 et 3

Table 2. Coding scheme and abbreviations

Coding dimensions	Explanations	Abbreviations
Citation / Location	Surname of first author and year of publication; Country where study undertaken.	* Studies reporting data for both SRH and CVD risk factors.
Health indicator / Analytic variable	Self-rated health: How would you describe your overall state of health: Excellent, very good, good, fair, or poor? Risk factors for CVD: Physical activity, diet, smoking, alcohol consumption, body mass index, overweight and obesity, diabetes, hypertension. Mortality: All-cause and cause specific mortality where individual-level data available. Analytic variable: Treatment of the outcome variable was treated.	BMI: Body mass index; CVD: Cardiovascular disease; CHD: Coronary heart disease; HBP: High blood pressure; N'hood: Neighbourhood; PA: Physical activity. Analytical variables: Dichotomous/binomial: 0/1; Ordinal/ordered categorical: ord; Continuous: cont.
Design / Year of data collection (individual-level)	Cross-sectional, longitudinal (follow-up), or case-control research design. Year of data collection at the individual level.	
Sample size individuals (sex/age range)	Sample size of individual data (full data set) and sex distribution and age in years (y) range of the sample.	y: year NR: Not reported
Individual characteristics adjusted for	Individual-level characteristics adjusted for in multilevel models.	A: Age; E: Education; ES: Employment status (e.g. employed, unemployed, retired); I: Income; MS: Marital status; N'hood: Neighbourhood; OS: Occupational status (type of work e.g. blue collar, professional); PA: Physical activity; R/E: Race/ Ethnicity; S: Sex; SC: Social class; SES: Socio-economic status; SN: Social network/support; All other characteristics are nominally identified.
Sample size of areas	Sample size and operational definition of areas	

Area-level exposures	Area-level exposure and type, i.e. whether they are derived/aggregated from individual-level data, e.g. census data, or integral that is only measurable at the area-level, e.g. number of parks.	[d]: Derived variable [i]: Integral variable N'hood: Neighbourhood
Crude between-area variation	Significant between-area variation unadjusted for individual-level characteristics unless otherwise specified. Reported by variance component and standard error, intraclass correlation coefficient for continuous and dichotomous outcomes, and plausible value range; all others are nominally identified. (Note: variance component is significant when > 1.96 * Standard error).	ICC: Intraclass correlation coefficient; p: p-value; Se: Standard error; PP: Predited probability; PVR: Plausible value range; VC: Variance component; NR: Not reported; Sign: Significant; NS: Not significant; SNR: Significance not reported
Adjusted between-area variation	Significant between-area variation adjusted for individual-level characteristics unless otherwise specified. Same as above	Same as above
Significant adjusted area effects	Significant area effects on health in models adjusting for individual and area-level variables (final models) unless otherwise specified. Cross-level interaction: Differential area effects across subgroup of individuals. In some cases, authors reported the association of an individual effect on a health indicator in a subset of areas. This was also considered to be a cross-level interaction. Area-level interaction: Differential effect of an area exposure on a health outcome conditional on another area exposure.	

Table 3. Results of study coding of multilevel investigations of area effects on self-rated health, cardiovascular morbidity and risk factors, and mortality among adults.

Citation /location	Health Outcome /Analytic variable	Design/year of data collection (individual level)	Individual sample size (sex/age range)	Individual characteristics adjusted for	Area sample size	Area-level exposures	Crude between area variation	Adjusted between area variation	Significant adjusted area effects
Self-rated health									
Béland (2002)[41] /Canada	Self-rated health (0/1)	Cross-sectional/ 1987	9422 66% Men ≥ 15y	A, S, E, I, ES, OS, MS, Stressful events, Perceived stress, SN, Social support, Locus of control, Sense of coherence	361 Contexts (Clusters of census tracts)	Unemployment rate [d], Gender distribution [d], Age group distribution [d], % Without high school diploma [d], % Immigrants [d], % Two-parent families [d], Average household income [d], Labour force participation rate [d], Occupational status [d].	VC (Se) = 0.049 (0.015)	NS (Adjusted also for area unemployment)	- Greater % of two-parent families, higher average income, and lower % of immigrants were associated with better SRH. - Cross-level interaction: Detrimental effect of deprivation on SRH was greater among those with greater perceived stress.
Blakely (2000)[42] /USA	Self-rated health (0/1)	Cross-sectional/ 1995; 1997	213695 % Men NR ≥ 15y	S, R/E, I Age-specific analyses	50 States	Median household income [d], Income inequality/Gini coefficient [i]. For 5 different time periods	NR	NR	- Greater inequality was associated with poorer SRH in the ≥ 45 y group. - In the ≥ 45 y group, the association between SRH and income inequality was stronger for earlier measures of inequality (1979-1981) than for concurrent measures (1995-1997).
Blakely (2001)[43] /USA	Self-rated health (0/1)	Cross-sectional/ 1995; 1997	279066 48% Men > 25y	A, S, R/E, I	50 States	Voting inequality [d], Voting turnout [d], Median Income [d], Income inequality/Gini coefficient [i].	NR	NR	- Greater voting inequality, lower voter turnout, higher income inequality, and lower median income were associated with poorer SRH. - Cross-level interaction: Effect of income inequality greater among ≥ 45y - Living in more equitable states was associated with better SRH among Whites, but only marginally among Blacks.

Blakely (2002)[44] /USA	Self-rated health (0/1)	Cross-sectional/ 1996; 1998	185479 48% men ≥ 0y	A, S, R/E, I	232 Metropolitan areas (MAs) 216 Counties 50 States	All area units: Income inequality/Gini coefficient [i], Average household income [d]. MAs: Cost of living index [d]. Counties: Population size [d].	NR	NR	- At the MA, county and state levels, lower income and higher income inequality were associated with poorer SRH, but in separate models only; adjusting for the cost of living index at the household level strengthened the association between MA average income and SRH, but did not change the inequality association. - Cross-level interaction: Association between state income inequality and poorer SRH was stronger among residents of rural areas.
Browning (2002)[45] /USA	Self-rated health (ord)	Cross-sectional/ 1997; 1999	2218 42% Men ≥ 18y	A, S, I, E, R/E, MS, Foreign born, Insurance coverage, Health behaviours, Health problems, Years of n'hood residency, Interview year	333 Neighbourhood clusters	Concentrated disadvantage [d], Residential stability [d], Immigrant concentration [d], Collective efficacy [d], Prior neighbourhood health [d], Violent victimization [d].	VC = 0.189 ICC =0.054 p<0.01	VC = 0.178 p<0.01	- Higher levels of collective efficacy associated with better SRH. - Cross-level interaction: The protective effect of individual education was greater in areas with higher collective efficacy.
Browning (2003)[46] /USA	Self-rated health (0/1)	Cross-sectional/ 1995; 1997; 1999	3272 41% Men ≥ 18y	A, S, I, E, R/E, MS, Foreign born, Insurance coverage, Years of n'hood residency, PA, Weight problems, Smoking, Interview year	339 Neighbourhood clusters	Affluence[d], Poverty [d], Residential stability [d], Immigrant concentration [d]. Hypothesised pathways - Collective efficacy[d], Social support and sociability [d], Subcultural tolerance for risk behaviour/anomie [d], Neighbourhood disorder [i], Prior n'hood health[d], Population size [d].	ICC=0.060 SNR	NS	- Greater affluence and lower levels of collective efficacy were associated with better SRH. - Greater residential stability and population size were associated with poorer SRH. - Area-level interaction: As area affluence increased, the positive effect of residential stability on poorer SRH decreased.

Browning (2003)[47] /USA	Self-rated health (0/1)	Cross-sectional/ Pooled survey data 1991-1999	8706 49% Men ≥ 18y	A, S, R/E, E, I, MS, Housing tenure, Interview year	342 Neighbourhood clusters	Affluence[d], Poverty[d], Residential stability[d], Immigrant concentration[d].	6.0% ($p < 0.001$)	NS (Adjusted also for area across time)	- Greater affluence was associated with better SRH.
Cagney (2005)[48] /USA	Self-rated health (ord)	Cross-sectional/ 1995; 1997; 1999	636 35% Men ≥ 55y	A, S, R/E, I, E, MS, Insurance coverage, PA, Weight problems, Interview year	246 Neighbourhood clusters	Age structure[d], Affluence[d], Poverty[d], Residential stability[d], Hypothesised pathway: Collective efficacy[d].	NR	NR	- Greater affluence was associated with lower odds of poorer SRH - Greater residential stability was associated with greater odds of reporting poorer SRH. - Area-level interaction: The detrimental effect of residential stability on SRH decreased as affluence increased.
Craig (2005)[49] /UK	Self-rated health (0/1)	Cross-sectional/ 1999/2000	18466 % Men NR 16-64y	A, S, E, ES	32 Local authorities	Mean household income[d], Income inequality /Gini coefficient[i] /Theil index[i], /90/10 ratio[i].	VC (Se) 0.015 (0.006)	VC (Se) 0.019 (0.007)	- Lower income was associated with poorer SRH.
Cummins (2005)[50] /UK	Self-rated health (0/1)	Cross-sectional/ Pooled data from two surveys 1: 1994-1999; 2: 1995; 1998	13899 45% Men ≥ 16y	A, S, SC, ES	425 Postcode sectors	Unemployment[d], Physical quality of residential environment[i], Public recreation[i], Crime[i], Access to multiple owned food shops[i], Access to banks and building societies[i], Health services[i], Left wing political climate[i], Political engagement[i] Access to private transport[i], Transport wealth[i].	NR	NR	- Poorer physical quality of residential environment, more left wing political climate, lower political engagement, higher unemployment, lower access to private transport, and lower transport wealth were associated with poorer SRH. - Cross-level interaction: Area effects for physical quality of residential environments, left wing political climate, access to private transport, and transport wealth on poorer SRH were greater among non-working individuals.
Drukker (2003)[51] /The Netherlands	Self-rated health (0/1)	Cross-sectional/ Year NR	3394 48% Men 20-65y	A, S, ES, Welfare recipient, Family type. Hypothesised	35 City-defined residential neighbourhoods	Composite index of socio-economic deprivation[d], Residential stability[d]. Hypothesised pathways - Housing conditions	NR	NR	- Greater deprivation was associated with poorer SRH. - When further adjusting for lifestyle, housing conditions, and perception of housing and neighbourhood contexts, the effect of deprivation on SRH was no

				pathway - Lifestyle: Smoking, PA, Drinking Fruit/vegetable intake, BMI		(person-bedroom index, residential type)[d], Individual perceptions of housing conditions[d], of area characteristics[d], of neighbourhood social contacts, maintenance, cosines, safety[d].			longer significant.
Drukker (2005)[52] /The Netherlands	Self-rated health (0/1)	Cross-sectional/ Year NR	3469 48% Men 20-65y	A, S, E, ES, Years of n'hood residency, Welfare recipient, Family type	35 City-defined residential neighbourhoods	Composite index of socio-economic deprivation[d], Residential stability[d].	NR	NR	- No significant main area effects. - Area-level interaction: In stable and very stable areas, higher deprivation was associated with poorer SRH
Franzini (2005)[53] /USA	Self-rated health (cont)	Cross-sectional/ 2001/02	3171 25% Men 18-94y	A, S, R/E, SES, Years of n'hood residency Hypothesised pathways – Physical activity; Social support.	100 Census block groups	Index of neighbourhood impoverishment[d], Hypothesised pathways – Collective efficacy[d], Physical and social disorder[i, d], Social processes relevant to children[d], N'hood racism[d], Social capital[d], N'hood climate of fear[d].	ICC = 0.17 Sign	ICC = 0.19 Sign	- Greater impoverishment was associated with poorer SRH. - Lower social capital, and greater levels of disorder and racism were associated with poorer SRH and mediated the effect of deprivation on SRH (which became non-significant).
Gee (2002)[54] /USA	Self-rated health (cont)	Cross-sectional/ 1993/94	1503 Chinese American % Men NR 18-65y	A, S, E, I, ES, SN, Medical insurance, Acculturation, Perceived racial discrimination	36 Census tracts	Poverty[d], Median housing value[i], Index of dissimilarity (greater segregation of Chinese)[d], Redlined areas (disfavoured home mortgage loan applicants)[i].	NR	NR	- Residing in redlined areas was associated with better SRH.
Gee (2004)[55] /USA	Self-rated health (cont)	Cross-sectional/ 1993/94	1503 Chinese American 47% Men 18-65y	A, S, E, I, ES, SN, Acculturation, Perceived traffic stress,	36 Census tracts	Poverty[d], Vehicular burden[d].	ICC = 0.03 SNR	NR	- Greater poverty was associated with poorer SRH. - Cross-level interaction: Greater vehicular burden was associated with poorer SRH among those perceiving

				Perceived environment					higher levels of traffic stress; adjusting for this cross-level interaction, effect of poverty on SRH no longer significant.
Hopman (2003)[56] /Canada	Self-rated health (NR)	Cross-sectional/ 1996/97	9423 31% Men ≥ 25y	A, S	9 Cities	None	NS	NS	- Examined between area variation only.
Hou (2003)[57] /Canada	Self-rated health (cont)	Cross-sectional/ 1996/97	8862 49% Men ≥ 12y	A, S, I, E, Smoking, PA, Emotional support	798 Census tracts	% Low-income[d], Income inequality/ Coefficient of variation[i].	VC = 0.042 ($p \leq 0.001$)	VC = 0.036 ($p \leq 0.001$)	- Higher % of low income associated with poorer SRH. - Higher income inequality associated with better SRH.
Hou (2005)[58] /Canada	Self-rated health (ord)	Cross-sectional/ 1996/97	34592 47% Men ≥ 12y	A, S, R/E, E, I, Immigrant status	487 Census tracts 25 Census metropolitan areas (CMAs)	Census tracts: Income inequality: /Gini index[i], /Median share[i], /Mean logarithmic[i], /Theil index[i], /Squared coefficient of variation[i]; Median income[d], % Adults[d], % University degree[d], % Seniors[d], % Single parent families[d], % Recent immigrants[d], % Non-white[d]. CMAs: Income inequality[i], Economic segregation[d], Median income[d].	NR	NR	- At the census tract level, higher income and % of people with a university degree were associated with better SRH; greater % of single parent family was associated with poorer SRH. - At the CMA level, within-CMA inequality positively associated with SRH.
Karlsen (2002)[59] /UK	Self-rated health (0/1)	Cross-sectional/ 1993/94	7848 % Men NR ≥ 16y	A, S, SC Race/ethnicity specific analyses	250 Electoral wards	Townsend deprivation index[d], Ethnic group density[d], Individual perceptions of problems of crime and nuisance[d], of lack of amenities[d], of environmental problems[d].	OR (95%CI) Whites: NS Caribbean: NS Indian: NS Pakistani/ Bangladeshi: 1.39 (1.16-1.67)	OR (95%CI) Whites: NS Caribbean: NS Indian: 1.17 (1.03-1.33) Pakistani/ Bangladeshi: 1.60 (1.25-2.03)	- Lower % of ethnic minority was associated with poorer SRH among Whites only.

Lindström (2004)[60] /Sweden	Self-rated health (0/1)	Cross-sectional/ 1994	3602 49% Men 20-80y	A, S, E, Foreign born, Social participation	75 City-defined administrative areas	None	VC (Se) = 0.096 (0.043) ICC = 2.8%	NS	- Examined between area variation only.
Malmström (1999)[61] /Sweden	Self-rated health (0/1)	Cross-sectional/ 1988/89	9240 50% Men 25-74y	A, S, E, BMI, Smoking, PA	837 Small-area market statistics	Care Need Index of deprivation[d], Townsend Deprivation Index[d].	NR	NR	- Greater deprivation *(both measures)* was associated with poorer SRH.
Mc Culloch (2001)[62] /UK	Self-rated health (0/1)	Cross-sectional/ Pooled survey data 1991-1999	10264 47% Men ≥ 16y	E, ES, R/E, Region of residence, Year of interview, Household type, Social housing, Access to car, Lag effect of antecedent SRH. Sex-specific analyses	634 Wards	Townsend index of deprivation[d]	NR	Women VC = 0.18 ($p< 0.001$) Men VC = 0.15 ($p< 0.001$) (Adjusted also for area deprivation)	- Greater deprivation was associated with poorer SRH. - Cross-level interaction: Effect of deprivation on SRH was greater among women residing in social housing.
Pampalon (1999)[63] /Canada	Self-rated health (0/1)	Cross Sectional/ 1992/93	20739 48% Men ≥ 15y 11439 households	A, S, E, I, ES, MS, Smoking, PA, Alcohol consumption	1833 Census enumeration areas 37 Regional subdivisions	None	NR	VC (Se) Enumeration areas = 0.066 (0.030) Regions: NS	- Examined between area variation only.
Patel (2003)[64] /USA	Self-rated health (ord)	Cross-sectional/ 1993/94	2561 Mexican American % Men NR ≥ 65y	A, S, E, I, MS, SN, Stressful life events, Obesity, Smoking, Medical conditions, Disability, Language acculturation, Financial strain	210 Census tracts	Economic disadvantage index[d], % Mexican Americans[d], Residence in a border USA/Mexico community[i].	ICC = 0.161 SNR	NR	- Higher economic disadvantage, lower % of Mexican Americans, and border community location were associated with poorer SRH. - Area-level interaction: Effect of area economic disadvantage on SRH was stronger among residents of border communities than among those of non-border communities.

Reijneveld (2000)[65] /The Netherlands	Self-rated health (0/1)	Cross Sectional/ Year NR	5121 % Men NR ≥ 16 y	A, S, I, E, ES	92 City-defined neighbourhoods 76 Postcode sectors 22 Boroughs	Mean income [d], % Benefit-dependent earners [d], % Low-income earners [d].	VC (Se) = N'hoods: 0.074 (0.027) Postcode sectors: 0.068 (0.025) Boroughs: 0.053 (0.024) (Age-sex adjusted)	NS	- For all three exposures, greater deprivation was associated with poorer SRH but only at the borough-level.
***Reijneveld (2002)**[39] /The Netherlands	Self-rated health (0/1)	Cross-sectional/ Pooled survey data 1991-2000	23269 % Men NR ≥ 16y	A, S, E	484 City-defined neighbourhoods 7 Cities	General practitioner (GP) deprivation score [d], Mean income [d], % Benefit-dependent earners [d].	NR	VC (Se) - Area (n'hood) GP deprivation score: 0.083 (0.034) Area mean income: 0.077 (0.034) Areas % social benefit: 0.076 (0.034)	- For all three exposures, greater deprivation was associated with poorer SRH at the neighbourhood level. - Area effects on SRH were not consistent across cities.
Robert (2002)[66] /USA	Self-rated health (NR)	Cross-sectional/ 1986	1669 41% Men ≥ 60y	A, S, E, I, R/E, Assets	48 States	Socio-economic disadvantage index [d]	VC (Se) = 0.10 (0.025) ICC = 0.14	VC (Se) = 0.04 (0.018)	- Greater socio-economic disadvantage was associated with poorer SRH.
Stafford (2001)[67] /UK	Self-rated health (ord)	Cross-sectional/ 1991-1993	6901 69% Men 35-55y	A, S, Employment grade, SN, Household deprivation, Perceived housing quality, Problems with	1831 Wards	Townsend Deprivation Index [d]	Women VC (Se) = 0.18 (0.08) Men VC (Se) = 0.30 (0.08) Women PP (PVR)=	Women : NS Men VC (Se) = 0.20 (0.07)	- Greater deprivation was associated with poorer SRH. - Cross-level interaction: Effect of deprivation on SRH was greater among those in lower employment grade.

				n'hood, Community participation			34% (18-54%) Men PP (PVR)= 20% (8-43%)		
***Stafford (2003)** [40] /UK	Self-rated health (0/1)	Cross-sectional/ 1997-1999	5539 70% Men Age NR	A, S, Employment grade, Problems with n'hood Financial problems, Dissatisfaction with standard of living, Position on social ladder	2112 Wards	Townsend Deprivation Index [d]	NR	NR	- Greater deprivation was associated with poorer SRH.
Stafford (2004) [68] /UK and Finland	Self-rated health (0/1)	Cross-sectional/ London: 1991-1993; Helsinki: 2000-2001	Pooled data; London 5301 61% Men 35-55y; Helsinki 4287 21% Men 40-60y	A, S, ES	England: 863 Census wards Finland: 223 City-defined neighbourhoods	Unemployment rate [d], % Single households [d], % Single parent households [d], % Manual workers [d].	VC (Se) London: 0.106 (0.040)/ ICC = 0.03 Helsinki: NS (Age-sex adjusted)	London: NS Helsinki: NS	- In pooled data, higher % of unemployment, % of manual workers, and % single households were associated with poorer SRH. - Higher % of single parent households was associated with poorer SRH in London but not in Helsinki.
Stafford (2005) [69] /UK	Self-rated health (0/1)	Cross-sectional/ Pooled data from two surveys 1994-1999; 1995/1998.	8437 45% Men ≥ 16y	A, SC, ES, family type Sex-specific analyses	238 Postal sectors	Individual perceptions of: family ties [d], friendship ties [d], participation [d], integration into the wider community [d], trust [d], attachment to n'hood [d], tolerance [d], being able to rely on others [d] Score of n'hood	VC (Se)/ % variation between area: Women: 0.212 (0.048)/ 6% Men: 0.141 (0.045)/ 4% (Age adjusted)	VC (Se) Women: 0.125 (0.039) Men: 0.005 (0.033)	- Women: Lower levels of integration into wider community, trust, tolerance, political engagement, access to banks and building societies, and access to private transport, fewer health services, poorer quality of environment, higher unemployment, stronger family ties and more left-wing political climate were associated with poorer SRH. - Men: Higher levels of crime and lower access to multiple food shops were associated with poorer SRH.

						Political engagement[i], Political climate[i], Crime[i], Access to multiple owned food shops[i], Access to banks and building societies[i], Health services[i], Public recreation[i], Quality of physical environment[i], Transport wealth[i], Access private transport[i], Unemployment[d].			
Subramanian (2001)[70] /USA	Self-rated health (0/1)	Cross Sectional/ 1993/94	144692 43% Men 18-98y	A, S, R/E, I, MS, Smoking, Health check-up, Health insurance coverage	39 States	Median income per capita[d] Income inequality/Gini coefficient[i], Social capital[d].	VC (Se) = 0.083 (0.020)	VC (Se) = 0.039 (0.010)	- Greater median income and social capital level were associated with better SRH. - Cross-level interaction: Greater income inequality was associated with better SRH among high income group only.
Subramanian (2002)[71] /USA	Self-rated health (0/1)	Cross-sectional/ 2000	21456 42% Men 18-89y	A, S, R/E, E, I, MS, Perceived individual trust	34 Communities	(d) Social trust (social capital)[d]	VC (Se) = 0.045 (0.015)	NS	- Greater level of social trust was associated with better SRH. - When adjusting for individual perceived trust, main effect of social capital no longer significant - Cross-level interaction: Protective effect of community social capital among high-trust individuals; among low-trust individuals, the effect was reversed.
Subramanian (2003)[72] /USA	Self-rated health (0/1)	Cross-sectional/ 1995; 1997	201221 47% Men ≥ 18y	A, S, R/E, MS, E, I, ES, Health insurance coverage	50 States	Income inequality/Gini coefficient[i], % Black residents[d].	VC (Se) = 0.031 (0.007) (Also adjusted for area income inequality)	VC (Se) = 0.029 (0.007) (Also adjusted for area income inequality)	- Greater income inequality associated with poorer SRH.
Subramanian (2005)[73] /USA	Self-rated health (0/1)	Cross-sectional/ 2000	21572 42% Men 18-89 y	A, S, R/E, MS, E, I	36 Communities	None	VC (Se) = 0.045 (0.014)	NS	- Examined between area variation only.

Subramanian (2005)[74] /USA	Self-rated health (0/1)	Cross-sectional/ 2000	51316 47% Men ≥ 18y	A, S, R/E, MS, E, I	207 Metropolitan areas	Black/white dissimilarity index[d], Black isolation index[d], White isolation index[d], Total population size[d], % Poor[d].	NR	VC (Se) = 0.037 (0.008) White/Black disparity in reporting poorer SRH 0.172 (0.48) such that: Whites 0.060 (0.012) Blacks: 0.090 (Se NR)	- Covariation between the metropolitan area variation in SRH for Whites and the White/Black disparity in poorer SRH: Greater White/ Black disparity was associated with better SRH among Whites. - Greater Black isolation index was associated with poorer SRH among Blacks only.
Veenstra (2005)[75] /Canada	Self-rated health (0/1)	Cross-sectional/ 2002	1435 53% Men ≥ 18y	A, S, E, I, Trust in community members, Political trust, Participation in voluntary associations.	25 Communities	Number public spaces per capita[i], Number of voluntary organisations per capita[i], Community and political trust[d], Median household income[d], Income inequality/ % of income held by poorest 50% households[i].	NS	NR	- No significant area effects.
Wen (2003)[76] /USA	Self-rated health (ord)	Cross-sectional/ 1996; 1997; 1999	3459 40% Men ≥ 18y	A, S, R/E, E, I, MS, Smoking, High blood pressure	343 Neighbourhood clusters	Affluence[d], Education level[d], Poverty[d], Income inequality/Gini coefficient[i], Prior n'hood health[d]. Hypothesised pathways: Physical disorder[i], Health enhancing services[i], Social resources[d], Crime[i].	NR	NR	- Higher affluence and education levels were associated with better SRH. - Potential pathways linking affluence to SRH included crime, social resources, and the physical environment.

Xi (2005) [77] /Canada	Self-rated health (0/1)	Cross-sectional/ 1996	30820 46% Men ≥ 25y	A, S, E, I, MS, PA, Smoking	Public Health Units	Median income [d], Income inequality/Gini coefficient [i].	VC (Se): 0.042 (0.012)	VC (Se): 0.020 (0.008)	- Lower median income and higher income inequality were associated with poorer SRH.

Citation /location	Health Outcome /Analytic variable	Design/year of data collection (individual level)	Individual sample size (sex/age range)	Individual characteristics adjusted for	Sample size of areas	Area-level exposures	Crude between area variation	Adjusted between area variation	Significant adjusted area effects
CVD and risk factors									
Chaix (2003)[78] /France	Overweight (0/1) Smoking (0/1) Drinking (0/1) Physical inactivity (0/1)	Cross-sectional/ 1999	12948 49% Men 16-75y	A, S, E, I, ES, MS, OS	95 Administrative departments	Gross domestic product (GDP) per capita [d], Type of county of residence (rural, small town, medium sized city, major city) [i].	NR	NR	- Higher GDP was associated with greater odds of being a dependent smoker among both women and men, and with being a dependent drinker among women only. - Cross-level interaction: Higher GDP was associated with overweight among blue-collar workers only.
Chaix (2004)[79] /France	Smoking (1/0; cont)	Cross-sectional/ 1999	12948 49% Men 16-75y	A, S, E, I, ES, MS, OS	95 Administrative departments	Gross domestic product (GDP) per capita [d], Type of county of residence (rural, small town, medium sized city, major city) [i].	NR	Smoking (0/1): VC (Se) = 0.014 (0.007)	- Higher GDP was associated with greater odds of smoking. - Among smokers, higher GDP was associated with greater levels of tobacco consumption.
Chang (2005)[80] /USA	BMI (cont)	Cross-sectional/ 1996-1998	143931 44% Men ≥ 18y	A, E, I Census region Sex-race specific analyses	226 Metropolitan statistical areas (MSAs)	Income inequality/Gini coefficient [i], Median household income [d], Population size [d].	VC = White women: 0.320 ($p<0.001$); White men: 0.174 ($p<0.001$); Black women: 0.780 ($p<0.001$); Black men: 0.473 ($p<0.001$)	NR	- Greater income inequality was associated with lower BMI among White women only. - Greater median income was associated with lower BMI among White women and men only.

Chuang (2005)[81] /USA	Smoking (cont)	Cross-sectional/ Pooled data from five surveys 1979-1990	8121 45% Men 25-74y	A, S, E, I, R/E	82 Census tracts and/or Census blocks	SES score[d], Within a one-mile radius around individual residence: Density of convenience stores[i], Distance of convenience stores[i], Number of convenience stores[i].	NR	NR	- Lower SES, higher convenience store density, and lower distance to convenience store were associated with greater odds of smoking, but in separate models only. - Cross-level interaction: In more affluent areas only, high-SES individuals were less likely to smoke. - Area-level interaction: In areas with higher convenience store density, the odds of smoking were similar across all income groups.
Cubbin (2005)[82] /USA	Behaviour change (0/1) 12-year probability experiencing CHD event (cont)	Cross-sectional/ Pooled data from five surveys 1979-1990	8197 45% Men 12-74y	A, S, R/E, MS, SES, City, Survey year	82 Census tracts and/or Census blocks	Townsend Index of deprivation[d].	NR	Significant, but no variance estimates reported (Also adjusted for area deprivation)	- Higher deprivation was associated with greater odds of no positive behaviour changes.
Diez-Roux (1999)[83] /USA	Dietary intake (cont): fruits, vegetables, meat, fish. Unhealthy dietary intake (0/1): fruits, vegetables, meat, fish. Nutrients intake (cont): saturated fat, polyunsaturated fat, cholesterol.	Cross-sectional/ 1987; 1989	13095 45% Men 45-64y	I, Residential location Sex-race specific analysis	Census block groups	Median household income[d]	NR	NS	- White men: Lower income was associated with lower intake of fruits and fish, and with greater odds of unhealthy dietary intake of fruits. - Black men: Lower income was associated with greater intake of cholesterol and polyunsaturated fat. - White women: Lower income was associated with lower intake of fruits and fish, greater intake of meat, and greater odds of unhealthy dietary intake of fish and meat. - Black women: Lower income was associated greater odds of unhealthy dietary intake of fruits.

Diez-Roux (2000)[84] /USA	Sedentarism (0/1) Smoking (0/1) BMI (cont) HBP (0/1)	Cross-sectional/ 1990	70534 43% Men ≥ 18y	I, R/E Sex-income specific analyses	44 States	Income inequality/Robin Hood Index[i]	NR	NR	- Among lower income women, greater inequality was associated with higher BMI and greater odds of HBP. - Among higher income women, greater inequality associated with greater odds of smoking. - Among both sex and across categories of individual-level income, greater inequality associated with greater likelihood of sedentary behaviour.
Duncan (1999)[85] /UK	Smoking (0/1)	Cross-sectional/ 1984/85	9003 % Men NR Age NR	A, S, E, SC, ES, MS, Housing tenure	396 Electoral wards 198 Regions	Ward-level composite index of deprivation[d]	NR	VC (Se) = Ward: 0.051 (0.02) Region: NS	- Greater deprivation was associated with greater odds of smoking.
Ecob (2000)[86] /UK	Smoking (1/0; cont) Alcohol consumption above safe limit (0/1) Alcohol consumption levels (cont) Diet (0/1) Exercise (0/1)	Cross-sectional/ 1987/88	15y:1009; 35y: 985; 55y:1042;	E, I, SC, MS, Moved in the 5 years preceding interview Sex-age specific analyses	Postcode sectors	Carstairs-Morris Index of deprivation[d]	NR	All age/both sex VC (Se) = Unhealthy diet 0.17 (0.07)	- Greater deprivation was associated with unhealthy diet lack of exercise, and smoking. - Cross-level interaction: Among lower income groups, deprivation was associated with unhealthy diet. - Among those aged 35y or 55y, greater deprivation was associated with unhealthy diet, lack of exercise, and smoking.
Ewing (2003)[87] /USA	Any PA (cont) Meeting guidelines for PA (0/1) Walking (cont) Obesity (0/1) BMI (cont) HBP (0/1) Diabetes (0/1) CHD (0/1)	Cross-sectional/ Pooled survey data 1998-2000	206992 % Men NR 18-75y	A, S, R/E, E, Smoking, Fruit/ vegetable consumption,	83 Metropolitan areas (MAs) 448 Counties	Metropolitan sprawl index[i], County sprawl index[i].	NR	NR	- Less sprawl at the county and MA levels were associated with higher levels of walking for leisure - Less sprawl at the county level was associated with lower BMI and lower odds of obesity and HBP.

Fisher (2004)[88] /USA	Walking: Latent variable comprising three items (cont): 1: Walk/ stroll in n'hood 2: Walk/any physical activity with neighbours 3: Walk/any physical activity in n'hood park	Cross-sectional/ 2001	582 31% Men ≥ 65y	A, S, R/E, MS, E, I, Health status, Walking efficacy	56 City-defined neighbourhoods	Social cohesion[d], N'hood problems[d], Safety[d], % Low income[d], % White residents Senior population[d], density[d], Facilities per n'hood acre[i].	ICC for the three items of walking - SNR: Item 1: ICC = 0.04 Item 2: ICC = 0.03, Item 3: ICC = 0.02.	NR	- Higher social cohesion, % of low-income households, % of senior residents, number of facilities for walking, and % of White residents, were associated with greater levels of walking.
Kairouz (2005)[89] /Canada	Drinking (1/0; cont), Alcohol Dependence (0/1)	Cross-sectional/ Year NR	4918 % Men NR ≥ 15y	A, S, MS, E, I	373 Census subdivisions	None	NS	NS	- Examined between area variation only.
Kavanagh (2005)[90] /Australia	Any physical activity (0/1) Walking (0/1) Cycling (0/1) Jogging (0/1) Swimming (0/1)	Cross-sectional/ Year NR	2349 44% Men 18-74y	A, S, E, I, OS	50 Collectors districts	% Low-income[d]	VC (Se) = Overall PA : 0.127 (0.043) Walking : 0.242 (0.079) Cycling : 0.226 (0.073) Jogging: 0.159 (0.058) Swimming : 0.142 (0.061) (Age-sex adjusted)	VC (Se) = Overall PA : NS Walking : 0.209 (0.071) Cycling : 0.201 (0.067) Jogging: NS Swimming : 0.127 (0.058)	- Higher % low-income was associated with lower levels of any physical activity and jogging.

Leyland (2005)[91] /UK	Presence of one or more CVD conditions (0/1)	Cross-sectional/ 1998/99	8804 43% Men 18-74y	A, S, SC, Smoking	312 Postcode sectors	Carstairs Index of deprivation [d], Social class [d], Smokers [d].	ICC = 0.011 SNR (Age-sex adjusted)	ICC = 0.010 SNR	- Greater deprivation was associated with greater odds of having one or more CVD conditions.
Li (2005)[92] /USA	Walking (cont)	Cross-sectional/ 2001	577 36% Men 65-94y	A, S, E, I, R/E, MS, Perceptions of: Proximity to recreational facilities, Safety for walking, Safety from traffic, Number of nearby recreational facilities	56 City-defined neighbourhoods	Density of places of employment [i], Density of households [i], Number of street intersections [i], Total green and open spaces for recreation [i]. Within 0.5 mile radius around the home: Number street intersections [i], Total green areas [i].	ICC = 0.28 SNR	NR	- Greater density of places of employment, density of household, number of street intersections, and number of green and open spaces were associated with higher levels of walking. - More favourable perception of number of recreational facilities and safety for walking were associated with higher levels of walking. - Cross-level interaction: In areas with greater street intersections, levels of walking were higher among those perceiving greater safety from traffic.
Lindström (2003)[93] /Sweden	Leisure-time physical activity (0/1)	Cross-sectional/ 1994	3377 49% Men 20-80y	A, S, E, Foreign born, Social participation	74 City-defined administrative areas	Residential mobility (social capital) [d]	VC (Se) = 0.171 (0.053)	NS	- No significant area effects.
Lindström (2003)[94] /Sweden	Smoking (0/1)	Cross-sectional/ 1994	3393 49% Men 20-80y	A, S, E, Foreign born, Social participation	77 City-defined administrative areas	None	VC (Se) = 0.085 (0.034) ICC = 0.025	NS	- Examined between area variation only.
Morland (2002)[95] /USA	Meeting dietary guidelines in (0/1): Fruits & Vegetables; Cholesterol; Total fat; Saturated fat	Cross-sectional/ 1993-1995	2392 Black Americans 36% Men 8231 White Americans 46% Men 35-74y	S, E, I Race/ethnicity specific analyses	208 Census tracts	Number of supermarkets [i], Number of small grocery stores [i], Number of full service restaurants [i], Number of fast-food restaurants [i].	NR	NR	- Black Americans: Greater number of supermarkets was associated with greater odds of meeting dietary guidelines for fruits and vegetable intake, total and saturated fat; greater number of full-service restaurants was associated with greater odds of meeting dietary guidelines for saturated fat. - White Americans: Greater number of supermarkets was associated with greater odds of meeting dietary requirements for total and saturated fat.

Morris (2001)[96] /UK	CHD incidence (0/1)	Follow-up/ 1978-1980; 1996	7735 Men only 40-59y	A, SC, Blood pressure, BMI, Serum total cholesterol, PA, Smoking, Alcohol consumption	24 Towns	Water hardness[i], Max temperature[i], Min temperature[i], Daily rainfall[i], Sunshine hours[i].	VC = 0.028 SNR (Age adjusted)	VC = 0.014 SNR	- No significant area effects.
***Reijneveld (2002)**[39] /The Netherlands	Smoking (0/1)	Cross-sectional/ Pooled survey data 1991-2000	23269 % Men NR ≥ 16y	A, S, E	484 City-defined neighbourhoods 7 cities	General practitioner (GP) deprivation score[d], Mean income[d], % Benefit-dependent earners[d].	NR	NS	- For all exposures, greater deprivation was associated with greater likelihood of smoking. - Area effects on smoking were not consistent across cities.
Robert (2004)[97] /USA	BMI (cont)	Cross-sectional/ 1986	3617 38% Men ≥ 25y	A, R/E, MS, E, I, ES, Assets, SN, PA, Smoking, Financial chronic stress, Number of stressful events Sex-specific analyses	48 States	Socio-economic disadvantage index[d], Income inequality/Gini coefficient[i], % Black[d].	Women: ICC = 0.06 (p<0.001) Men: ICC = 0.11 (p<0.001)	VC (Se) = Women: 1.05 (0.33) (p<0.01) Men: 1.74 (0.48) (p<0.001)	- Higher socio-economic disadvantage and income inequality were associated with higher BMI among women only.
Scribner (2000)[98] /USA	Alcohol consumption (cont)	Cross-sectional/ Year NR	2604 % Men NR > 18y	A, S, R/E, E	24 Census tracts	Mean distance to the closest alcohol outlet[i]	11.5% SNR	NR	- Greater mean distance to closest alcohol outlet was associated with lower alcohol consumption.
***Stafford (2003)**[40] /UK	Wais to hip ratio (cont)	Cross-sectional/ 1997-1999	5539 70% Men Age NR	A, S, Employment grade, Problems with n'hood, Financial problems, Dissatisfaction with living standard, Position on social ladder	2 112 Wards	Townsend Deprivation Index[d]	NR	NR	- Higher deprivation was associated with higher mean waist/hip ratio.

Stjärne (2004) [99] /Sweden	Myocardial infarction (MI) (rate)	Case-control/ 1992-1994	Case: 1546 69% Men Control: 2064 66% Men 45-70y	A, E, OS, Labour market position, Cohabiting/non cohabiting, SN, Smoking, BMI, Hypertension, Sex-specific analyses	862 Small residential areas	Townsend Deprivation Index [d], Congdon's Index of social fragmentation [d].	Women: VC = 0.146 SNR Men: VC = 0.065 SNR	NR	- Greater deprivation and social fragmentation were associated with greater risk of MI; the effect was stronger among women than among men.
Sundquist (1999) [100] /Sweden	Obesity (0/1) Smoking (0/1) PA(0/1)	Cross-sectional/ 1988/89	9240 % Men NR 25-74y	A, S, E	8519 Small-area market statistics	Care Need Deprivation Index [d], Townsend Deprivation Index [d].	NR	NR	- Greater deprivation (as measured by both indices) was associated with greater likelihood of smoking, physical inactivity, and obesity.
Sundquist (2004) [101] /Sweden	Coronary heart disease (rate)	Follow-up/ 1995-1999	2637628 50% Men 40-64y	A, I Sex-specific analyses	8547 Small-area market statistics	Care Need Deprivation Index [d]	NR	Women: VC (Se) = 0.084 (0.009) ICC = 0.025 Men: VC (Se) = 0.034 (0.004) ICC = 0.010	- Greater deprivation was associated with higher risk of developing coronary heart disease.
Sundquist (2004) [102] /Sweden	Coronary heart disease (rate)	Follow-up/ 1986-1993; 1997	25319 49% Men 35-74y	A, S, E, I, Smoking, Years of n'hood residency	6145 Small-area market statistics	% Less than 10y of education [d], % Income in lowest national quartile [d].	NR	VC (Se) = 0.25 (0.10) (Also adjusted for area income and education)	- Lower education and income levels were associated with higher risk of developing coronary heart disease.
Tonne (2005) [103] /USA	Survival after acute myocardial infarction (AMI) (cont)	Cross-sectional/ 1995; 1997; 1999; 2001	3423 58% Men ≥ 25y	A, S, R/E, Medical history; Clinical complications; AMI order and type	111 Census tracts	Median income [d], % Below poverty line [d], % Low education [d], % Overcrowding [d], Composite deprivation score (four measures) [d].	NR	NR	- Greater levels of overcrowding, % living below the poverty line, % low education, greater deprivation, and lower median income were associated with lower survival after AMI.

Twigg (2000) [104] /UK	Smoking (0/1) Alcohol consumption (0/1)	Cross-sectional/ Year NR	Sample size NR % Men NR ≥ 16y	A, S, MS	Wards	% Population with no car [d], % Population with dual car ownership [d], % Population privately renting [d], % Population higher social class (I or IIa) [d].	NR	NR	- Higher % of population with no car and privately renting associated with greater odds of smoking; higher % of dual car ownership was associated with lower odds of smoking. - Higher % of population of higher social class, privately renting, and dual car ownership were associated with greater odds of problem drinking. - Cross-level interactions: Single women living in areas of higher % of privately rented households had higher odds of problem drinking; those living in more affluent areas had higher odds of smoking.
Van Lenthe (2002) [105] /The Netherlands	Overweight (0/1)	Cross-sectional/ 1991	8897 49% Men 20-70y	A, S, E	86 Administrative neighbourhoods	Index of neighbourhood deprivation [d]	NR	NS	- Greater deprivation was associated with higher odds of overweight. - Cross-level interactions: Higher deprivation was associated with greater odds of overweight among women, in the ≥ 49 y group, and in all education groups except highest.
van Lenthe (2005) [106] /The Netherlands	Walking/ cycling to shops/work (0/1); Walking/ cycling/ gardening in leisure time (0/1); Sport participation (0/1)	Cross-sectional/ 1991	8767 49% Men 20-70 y	A, S, E	78 Administrative neighbourhoods	Index of neighbourhood deprivation [d] General physical design of neighbourhood [i], Quality green facilities [i], Noise pollution from traffic [i], Police attention required [i], Availability food shops [i], Availability sports and recreational facilities [i].	NR	NR	- Greater deprivation was associated with lower odds of almost never walking/ cycling to shops or work, with greater odds of almost never walking, cycling or gardening in leisure time, and with lower odds of sport participation. - Poorer general physical design and higher levels of noise pollution were associated with greater odds of almost never walking/cycling and gardening during leisure time. - Less proximity to sports facilities and low levels of police attention associated with low participation in sport. - Cross-level interactions: Odds of almost never walking/cycling to shops or work lower in the 20-49 y group in areas with greater noise pollution; greater in the

								50-70 y group in areas with less proximity to shops; and lower in the 50-70 y group in more deprived areas.
Wendel-Vos (2004) [107] /The Netherlands	Walking/ cycling for (cont): Recreation; Commuting; Overall.	Cross-sectional/ 1998	11541 46% Men 20-59 y	A, S, E	300m radius and 500m radius around postal code	Square area of green and recreational space: Woods [i], Parks [i], Sport grounds [i], Allotments [i], Day-trip grounds [i].	Between person variance 100 times greater than between postal code variance; SNR	NR	- In areas defined by the 300m radius only, square area of sport ground was associated with more overall cycling, cycling for leisure and for commuting purposes; square area of parks was also significantly associated with more cycling for commuting.

Mortality

Citation /location	Health Outcome /Analytic variable	Design/year of data collection (individual level)	Individual sample size (sex/age range)	Individual characteristics adjusted for	Sample size of areas	Area-level exposures	Crude between area variation	Adjusted between area variation	Significant adjusted area effects
Borrel (2002)[108] /Spain	Mortality all injury (rate); Mortality from (rate): Traffic injury; Falls; Drug overdose; Suicide; Other injuries.	Cross-sectional/ 1992-1998	4393 63% Men > 19y	E Age-sex specific analyses	38 City-defined neighbourhoods	% Unemployed men [d], % Men in jail [d].	NR	Mortality from all injury Men: 0.016 (p=0.001) Women: 0.027 (p<0.001) (also adjusted for % unemployed men); Mortality from drug overdose Women: 0.295 (p<0.001) (also adjusted for % men in jail)	- Higher % of unemployed men was associated with overall injury mortality, and with mortality from falls among men only. - Higher % of men in jail was associated with greater likelihood of fatal drug overdose.
Bosma (2001)[109] /The Netherlands	All-cause mortality (rate)	Cross-sectional/ 1991	8506 % Men NR 15-74y	A, S, E, OS, Health status, Being unemployed or disabled, Severe financial problems	86 Administrative neighbourhoods	% Primary schooling [d], % Unskilled manual workers [d], % Unemployed/disabled [d], % Severe financial problems [d].	Significant, but no variance estimates reported.	NS.	- Greater % of unemployment/disability and presence of severe financial problems were associated with greater all-cause mortality.

Curtis (2004)[110] /UK	All-cause mortality (rate)	Longitudinal /1939; 1981	62719 40% Men 40-59y	A, S, SC, MS, ES, Housing tenure	192 Categories of residential areas	1981 Ward of residence: Carstairs Index of deprivation [d], Broad regional location [d]. 1939 Area of residence: In economic depression [d], Population density [d], % Population semi-skilled or unskilled manual work[d] % Population in overcrowded housing [d], Unemployment rate [d].	NR	NR	- Mortality risks were lower for people originating from affluent areas in 1939 after controlling for 1981 area type.
Franzini (2003)[111] /USA	Years of life lost (YLL) to premature CVD mortality (cont)	Cross Sectional/ 1991	50268 % Men NR ≥ 25y	S, R/E, Age adjusted education	12344 Census block groups; 3788 Census tracts; 247 Counties	Census block groups: Median house value [d] Census tracts: % College degree [d], % Blacks [d], % Hispanics [d], % Homeownership [d]. Counties: % High school diploma [d], % Some college [d], % College degree [d], Median income [d], Poverty rate [d], Unemployment rate [d], % Blacks [d], % Hispanics [d], % Homeownership [d], Crime index [d], Income inequality /Robin hood index [i], /90/10 ratio [i].	NR	Block groups: NS Census tracts: 3.10 ($p<0.05$) County: 0.46 ($p<0.05$) (Also adjusted for block group, census tract, and county variables).	- Block-group level: Higher median house value was associated with reduced YLL. - Census tract level: Greater % Blacks and % Hispanics were associated with decrease in YLL for Blacks and Hispanics. - County level: Higher % of homeownership was associated with lower YLL, and worse crime index was associated with increased YLL. - Cross-level interactions: Effect of lower crime at the county level and YLL was weaker among women. - Effect of median house value on YLL was greater among high-education individuals. - Effect of county level homeownership on YLL was stronger for Blacks. - Effect of median house value of census block groups on YLL more important in counties with higher % of homeownership.

Hembree (2005)[112] /USA	Overdose mortality (ratio)	Case-control/ 1996	1178 77% Men 15-64y	A, S, R/E	59 Residential community districts	(d) Median household income[d], (d) N'hood drug use[d]. External built environment - % buildings in: Dilapidated condition[i], Deteriorating condition[i], External wall problems[i], Window problems[i], Stairway problems[i], Clean streets/sidewalks[i], Structural fires[i]. Internal built environment - % housing units with: Toilet breakdowns[i], Non-functioning kitchen facilities[i], >3 Heat breakdowns in winter[i], Additional heating needs in winter[i], Large area of peeling plaster/paint[i], Internal water leakage[i].	NR	NR	- All characteristics of the external built environment, except % buildings with any external wall problems, were associated with a higher likelihood of overdose mortality. - Greater % of housing units experiencing toilet breakdowns, needing additional heating in winter, with large areas of peeling plaster or paint were associated with higher likelihood of overdose mortality.
Jaffe (2005)[113] /Israel	All-cause (rate); CVD mortality (rate); Non-CVD mortality (rate)	Follow-up 1983-1992	141683 45% Men 45-89y	MS, E, Origin, Continent of origin Age-sex specific analyses.	882 Statistical areas	Religious affiliation[d], SES index[d].	NR	VC (Se): Women: NS Men all ages: 0.020 (0.005) (Also adjusted for area variables)	- Greater deprivation was associated with greater risk of mortality. - Adjusting for area SES, stronger religious affiliation was associated with lower risk of all-cause mortality and CVD mortality among men and women, and with lower risk of non-CVD mortality among men only. - Area-level interaction: In lower and average SES areas, religious affiliation was protective of mortality among women only.

Jaffe 2005)[114] /Israel	All-cause mortality (rate)	Follow-up/ 1983-1992	131156 45% Men 45-89y	MS, E, I, Origin, Number of rooms in the house, Household amenities score Age-sex specific analyses	882 Statistical areas	SES index [d]	NR	VC (Se) Women 45-69y: 0.023 (0.010)	- Greater deprivation was associated with greater risk of mortality.. - Cross-level interaction: Effect of deprivation on mortality risks were lower among low-income men in the 45-69 y group, and higher among high-income men in the 70-89 y group.
Jerrett (2005)[115] /USA	All cause mortality (rate) Mortality by (rate): Ischemic heart disease (IHD); Cardio-pulmonary; Lung, digestive, ther cancers; Endocrine; Diabetes; Digestive; Accidents; All others	Follow-up/ 1982-2000	22905 % Men NR Age NR	44 individual variables: demographics, lifestyle, dietary, occupation, education	267 Zip Code areas	Income [d], Income inequality [d], Education [d], Population size [d], Racial composition [d], Unemployment [d], Potential exposure misclassification [i], % Air conditioning [i], Particulate matter (PM_{25}) [i], Ozone (O_3) [i], Intersection with highway [i].	NR	NR	- Greater concentration levels of PM_{25} and O_3, and presence of highway intersection were associated with greater risk of all-cause mortality, and with mortality from IHD, cardiopulmonary diseases, and with diseases of the endocrine and digestive systems. - When further adjusting for area SES, social factors, and % air conditioning, the effect of air pollution on all-cause and cause-specific mortality was no longer significant.
Lochner (2003)[116] /USA	All cause mortality (rate); Mortality by (rate): Heart disease; Cancer; Other causes	Cross-sectional/ 1994-1996	Sample size NR 68% Men 45-64y	A Sex-race specific analyses	342 Neighbourhood clusters	% Perceived reciprocity [d], % Perceived trust [d], Associational membership per capita [d], Composite index of economic disadvantage [d].	NR	NR	- Greater economic disadvantage was associated with higher risk of all-cause mortality and mortality from heart disease, and with mortality from "other causes" except among Black men. - Higher levels of reciprocity, trust, and civic participation were associated with lower risk of all-cause mortality and mortality from "other causes", and with lower mortality from heart disease among Whites only.

Marinacci (2004) [117] /Italy	All cause mortality (rate); Mortality by (rate): Diabetes; Stomach cancer; Lung cancer; Psychol discomfort; CHD; Cerebro-vascular diseases; Respiratory diseases.	Longitudinal /1971-1980; 1981-1991; 1991-1999.	1971/80: 799564; 1981/91: 889432; 1991/99: 821736. % Men NR $\geq$ 15y	E, Place of birth, Composite index of housing conditions Sex-age specific analyses	23 City-defined neighbourhoods	Neighbourhood Deprivation index [d]	NR	NR	- Greater deprivation was associated with higher risk of mortality. - Cross-level interactions: Effect of deprivation on mortality by coronary heart and respiratory diseases was greater among women and men aged 15 to 64 years; deprivation increased risk of mortality by cerebrovascular diseases but only among women.
Martikainen (2003) [118] /Finland	All cause mortality (rate); Mortality by (rate): Lung cancer; Other cancers; Disease of circulatory system; Other diseases; Accidents/ violence; Alcohol-related mortality	Follow-up/ 1991-1995	251509 Men only $\geq$ 25y	E, SC, Housing tenure, MS, Housing density Age-specific analyses	55 City-defined neighbourhoods	% Manual workers [d], % People aged > 60y [d], Social cohesion [d].	Average relative deviation i.e. how many % on average does the mortality rate of an area differs from total mortality rate: 25-64y = 16.6% Over 64y = 8.1 % SNR (Age adjusted)	Average relative deviation 25-64y = 4.1% Over 64y = 2.6% SNR	- Greater % of manual workers was associated with higher risk of all-cause mortality, higher risk of mortality from CVD, accident/violence, and alcohol in the 25-64y group, and with higher risk of mortality from CVD in the $\geq$ 65y group. - Lower social cohesion was associated with higher risk of all cause mortality and mortality from accident/violence and alcohol in the 25-64y group only. - Cross-level interaction: In the $\geq$ 65y group, those with lower education had higher risk of mortality in areas with both high and low % of manual workers and in areas with high social cohesion.

Mohan (2005) [119] /UK	Overall mortality (rate)	Follow-up/ 1984/85-2001	7578 % Men NR Age NR	A, S, SC, Housing tenure, Smoking, PA, Diet, Social capital (feel part of community, people to rely upon, feel lonely)	396 Electoral wards; 198 Parliamentary constituencies	Carstairs deprivation index [d], % In voluntary activity [d], % Core volunteers [d], % Social activity [d], % Altruistic activity [d]; % Political activity [d], % Voters in last election [d], % Thinking local friends are important [d], % Belonging to n'hood [d], % Would work to improve n'hood [d], % Talking to neighbours [d] % Feeling the local area is friendly [d], Standardised blood donation rate [d].	NR	NR	- Greater deprivation and lower % of persons involved in voluntary, core volunteering, social, altruistic, and political activities were associated with higher risk of mortality, but in separate models only.
Roos (2004) [120] /Canada	Overall mortality (rate)	Follow-up/ Nova Scotia: 1990-1999; Manitoba: 1996/97-2002	Nova Scotia 2116 48% Men 18-75y Manitoba 8032 47% Men 18-75y	A, S, I, E, Smoking, Diabetes, BMI, Residential mobility	Census enumeration areas	Household income [d], Dwelling value [i], % Education < grade 9 [d], Unemployment rate [d], % Single mother [d].	NR	NR	- No significant main area effects. - Cross-level interaction: For Manitoba only, lower income individuals had greater mortality risk in more affluent areas than in less affluent areas.
Subramanian (2005) [121] /USA	Overall mortality (rate)	Cross-sectional/ 1989-1991	79813 cells (6016425 individuals grouped in 79813 cells cross-tabulated by age, sex, race/ ethnicity) % Men NR ≥ 0y	A, S, R/E	5532 Census block groups 1307 Census tracts	% Population living below poverty line in the census tracts [d]	NR	Census tracts: Blacks VC = 0.524 ($p < 0.001$) Whites VC = 0.085 ($p < 0.001$)	- Greater poverty was associated with higher mortality risk especially among Blacks. - Poverty accounted for racial/ ethnic-specific heterogeneity in mortality at the census tract level.

Veugelers (2001) [122] /Canada	Overall mortality (rate)	Follow-up/ 1990-1999	2116 48% Men 18-75y	A, S, E, I, Smoking, Diabetes, BMI	705 Census enumeration areas	Household income [d], Dwelling value [i], % Education < grade 9 [d], Unemployment rate [d], % Single mother [d]	NR	NR	- No significant main area effects. - Cross-level interaction: Within affluent area only, mortality risks were lower among high-income individuals.
Waitzman (1999) [123] /USA	Overall mortality (rate)	Follow-up/ 1986-1994; 1995	136956 % Men NR 35-65y	A, R/E, E, I	34 Metropolitan areas	Spatial inequality/ economic segregation [d], Income inequality/Gini coefficient [i], Median income [d].	NR	NR	- Greater economic segregation and income inequality were associated with greater risk of mortality.
Yen (1999) [124] /USA	Overall mortality (rate)	Follow-up/ 1983; 1994	996 43% Men 36-96 y	A, S, R/E, Smoking, BMI, Alcohol consumption, Perceived health status	Census tracts	Social environment score: Population SES: Per capita income [d], % White-collar employees [d], % Crowding [d]; Commercial stores per 1000 people [i]; Environment/housing: Population size [d], Area (square miles) [i], % Renters [d], % Single-family dwellings [d].	NR	NR	- Poorer social environment was associated with higher risk of mortality. - Lower population SES, lower environment/housing score, and more commercial stores were associated with higher risk of mortality. - Cross-level interactions: Lower environment/housing score and higher population SES were associated with greater risk of mortality among lower income individuals.

Annexe II

Méthodologie : Notes supplémentaires

MÉTHODOLOGIE (notes supplémentaires)

Potentiel de vie active des milieux résidentiels

Choix des indicateurs

Dans la littérature sur les déterminants environnementaux de la marche, le choix des indicateurs du potentiel de vie active des milieux, de même que leurs mesures, varient selon les études. Les indicateurs les plus fréquemment utilisés sont liés à la mixité des usages du sol, à l'accessibilité aux services, à la densité de la population ou densité résidentielle, à la connexité et configuration du réseau de rues (Transportation Research Board and Institute of Medicine of the National Academies [TRB & IOM], 2005). Chacun de ces indicateurs sous-tend des processus d'influence différents sur les comportements de marche (Leslie, Coffee, Frank, Owen, Bauman, & Hugo, 2007).

Les chercheurs proposent que la densité de population influence positivement les comportements de marche de par la masse critique qu'elle crée dans les milieux de vie: plus il y a de gens qui marchent, plus il y a de possibilités de voir des gens marcher et de se sentir en sécurité et donc plus il y a d'incitatifs pour la marche (Forsyth, Oakes, Schmitz, & Hearst, 2007). Une plus grande densité de population implique aussi une plus grande congestion automobile au point où marcher, ou marcher pour prendre le transport en commun, est plus facile que de se déplacer en automobile. Une plus grande mixité des usages du sol confère la norme d'une diversité importante dans la trame urbaine qui favorise la marche vers les diverses destinations. L'accessibilité à des services multiples et diversifiés offre l'opportunité de se déplacer plus fréquemment pour les achats sur une distance réduite, et donc pouvant être parcourue à pieds. Finalement, la connexité du réseau de rues procure un plus grand choix de routes possibles pour se rendre à une destination, mais aussi des temps de déplacements plus courts.

Bien que la densité de la population, la mixité de l'occupation du sol, l'accessibilité aux services et la connexité du réseau de rues réfèrent à des dimensions uniques et à des processus d'influence multiples sur les comportements de marche, ces indicateurs du potentiel de vie active du milieu se concrétisent souvent aux mêmes endroits dans l'environnement urbain. L'environnement bâti émerge de l'interaction de trois composantes qui encouragent ou restreignent les activités de marche : la configuration de l'occupation du sol, le système de transport et le design urbain (TRB & IOM, 2005). Dans le cadre de la thèse, le choix a été fait d'étudier les caractéristiques des milieux résidentiels liées à la configuration des usages du sol et ayant le potentiel de favoriser une vie active parmi les populations locales. Trois indicateurs de la configuration des usages du sol ont été retenus, la densité de population, la mixité des occupations du sol et l'accessibilité à des services de proximité.

De futures études pourraient porter sur d'autres indicateurs, par exemple la connexité du réseau de rues, la structure du réseau piétonnier (trottoirs, traverses piétonnes, etc.), la vitesse de circulation automobile et l'attrait esthétique (indice de végétation, style architectural).

Mesures du potentiel de vie active calculées au niveau des aires de diffusion (AD)

Densité de population. La densité de la population réfère au nombre d'individus par unité de surface, ici au ratio entre le nombre total d'habitants et la superficie (km^2) de l'AD. Les données sur la taille de la population pour chacune des AD proviennent du fichier du recensement canadien de 2001.

Cette mesure de densité de population est brute, c'est-à-dire qu'elle tient compte de toute l'unité de surface de l'AD et non pas uniquement de la surface résidentielle (dans ce cas, il est question de densité nette). Une limite méthodologique d'une mesure brute, est que plus l'aire de l'AD est importante, plus la densité de population diminue en raison d'occupations non-résidentielles du sol ce qui a pour effet de diluer l'estimation totale de la densité de population. Des mesures plus précises de densité de la population par surface résidentielle ou encore de densité des logements par superficie développée auraient pu être calculées.

Néanmoins, une mesure brute de la densité de la population a été retenue, et ce, pour trois raisons. D'abord, puisque la marche se pratique à la fois dans les endroits résidentiels et non-résidentiels, une mesure brute de densité représente donc la densité de la population dans son ensemble dans une aire donnée, et est donc une mesure du potentiel de vie active du milieu appropriée (Forsyth et al., 2007). Ensuite, les diverses mesures de densité de population sont fortement associées les unes aux autres sur l'Île de Montréal. Par exemple la corrélation (Pearson) entre la densité de population et la densité de logements par km^2 est de 0.96 ($p<0.001$), alors que la corrélation (Spearman) entre la densité de population et la densité de population ajustée selon la superficie résidentielle de l'AD est de 0.83 ($p<0.001$). Ces fortes corrélations indiquent que le choix d'une mesure n'introduit pas d'erreur significative en comparaison à d'autres mesures. Finalement, il se peut que, pour une même densité de logements, la densité de population soit différente.

Quartier	**Densité de logements**	**Densité de population**
Petit Plateau	6646	11 223
Mile End	6927	13 603
Parc-Extension	7429	18 945

C'est le cas par exemple en comparant les quartiers de Parc Extension, Plateau Mont-Royal et Mile End, qui bien qu'ayant des densités de logement similaires, la densité de la population est plus élevée dans Parc Extension que dans les deux autres secteurs.

<u>Mixité des occupations du sol.</u> La mixité des occupations du sol correspond à la diversité des occupations du sol par unité de surface. Elle est été calculée à l'aide d'un indice d'entropie (Theil, 1972; Theil, 1971) soit :

$$E_j = -\sum_{i=1}^{n} \left[\left(A_{ij} / D_j \right) \ln \left(A_{ij} / D_j \right) \right] / \ln n \qquad \text{(Équation 1)}$$

Où A_{ij} est la superficie de l'occupation du sol i dans l'aire de diffusion j, D_j est la superficie de l'aire de diffusion j, et n est le nombre total d'occupations du sol. L'indice d'entropie varie entre 0 et 1, où 1 correspond à une AD diversifiée, et 0 à une AD caractérisée par une seule occupation du sol. Cet indice a été utilisé dans plusieurs études pour caractériser la diversité des occupations du sol (Frank, Schmid, Sallis, Chapman, & Saelens; Cloutier, Apparicio, & Thouez, 2007).

La mixité des occupations du sol a été calculée à partir du plan de l'occupation du sol de la Ville de Montréal, dans lequel le territoire est catégorisé en 16 occupations (Communauté urbaine de Montréal, 2001)[1]. L'occupation du sol réfère à « l'utilisation ou la fonction réelle et actuelle de chaque terrain sur le territoire » (Service de la mise en valeur du territoire; Division de l'aménagement, 2001). Les occupations du sol de chacun des polygones ont été vérifiées à l'aide de photos aériennes et de visites sur le terrain et validées par les données du rôle de d'évaluation de 1999.

Il aurait été possible de mesurer la mixité des usages du sol à l'aide d'autres sources de données, telles la carte cadastrale et la carte du rôle foncier, mais ces dernières n'ont pas été retenues. D'abord, la carte cadastrale n'est pas diffusée par la ville de Montréal; il était donc impossible de travailler avec cette dernière. La carte du rôle foncier, bien qu'en certains cas plus précise que le plan des occupations du sol, contient de nombreuses limites. D'abord, les seules occupations retenues sont les occupations résidentielles, commerciales et surfaces de bureaux et donc n'offrent pas

[1] Les 16 occupations du sol sont : Habitation faible densité; habitation moyenne densité; habitation haute densité; commerce de détail; centre commercial; édifice à bureaux; équipement et service communautaire; service d'utilité publique; industrie; carrière; site d'enfouissement; espace vert; golf; cimetière; espace rural; espace vacant.

un aperçu complet des différentes occupations du sol sur un territoire donné. En somme, cette base de données bien que très intéressante, est difficile à exploiter.

Comme une plus grande mixité des usages du sol confère l'idée d'une diversité dans l'environnement qui favorise la marche vers diverses destinations (Leslie et al., 2007), les 16 occupations du sol différentes ont été retenues pour le calcul de l'indice d'entropie. En cela, la mesure est similaire à celle utilisée par certains groupes d'auteurs (Cerin, Leslie, du Toit, Owen, Frank, 2007) mais diffère de celle utilisée par d'autres chercheurs dans des études où la mixité des usages du sol est fonction de 3 occupations, c.-à-d. résidentielle, commerciale et surface de bureaux (Frank et al., 2005). Cet indice ne détermine pas le type d'occupation du sol dominant dans l'aire d'étude, mais bien jusqu'à quel point l'occupation du sol est homogène, c.-à-d. si elle est caractérisée essentiellement par une fonction, ou si diverses fonctions sont combinées.

Dans des études futures, il serait intéressant d'explorer la création de typologies des occupations du sol (déterminée à l'aide du K-means par exemple). Une telle typologie permettrait de déterminer, par exemple, les AD caractérisées par des occupations résidentielles et commerciales, ou encore par des types d'occupations industrielles et terrains vacants.

Accessibilité géographique aux services de proximité. L'accessibilité géographique réfère à la facilité avec laquelle il est possible de rejoindre les services et les ressources (Hewko, Smoyer-Tomic, & Hodgson, 2002). Les mesures d'accessibilité les plus couramment utilisées sont la distance au service le plus proche, le nombre de services compris dans un rayon de n mètres, la distance moyenne à l'ensemble des services ou aux n services les plus proches et le modèle gravitaire. Chacune de ces mesures réfère à une conceptualisation différente de l'accessibilité soit (respectivement) la proximité immédiate, l'offre dans une unité géographique donnée, l'offre de services dans l'environnement immédiat, le coût moyen à partir d'une origine pour rejoindre les services ou pour rejoindre une offre diversifiée de services et l'attractivité potentielle des services en fonction de leur taille à partir d'une origine donnée (Apparicio, Abdelmajid, Riva, & Shearmur, 2008).

Plusieurs études ont démontré que le choix de la mesure d'accessibilité importe car chacune produira des cartes d'accessibilité différentes (Apparicio, et al., 2007; Smoyer-Tomic, Hewko, & Hodgson 2004). Dans la thèse, l'accessibilité aux services a été mesurée en termes d'offre de services dans un environnement immédiat, puisque parmi les mesures décrites, elle correspond plus étroitement à l'opportunité de se déplacer plus fréquemment pour atteindre certains services situés à distance de marche d'un point d'origine. Cette distance a été fixée à 1000 mètres en fonction du réseau de rues.

Jusqu'à présent, bien que certaines études démontrent qu'une plus grande accessibilité aux services et infrastructures est associée à des quantités et probabilités de marche plus importantes (Cerin et al., 2007; Humpel, Owen, Leslie, Marshall, Bauman, & Sallis, 2004; Lovasi, Vernez-Moudon, Pearson, Hurvitz, Larson, Siscovick, et al., 2008; Owen, Humpel, Leslie, Bauman, & Sallis 2004), il y a peu de consensus sur les services à inclure dans les calculs d'accessibilité. En lien avec le concept de vie active, l'hypothèse sous-jacente de l'influence de l'accessibilité des services et la marche, et que les services et ressources de consommation de « la vie de tous les jours » donc ceux utilisés sur une base régulière, puissent être accessibles à pieds. Dans cet optique, et en lien avec la conceptualisation de l'accessibilité définie selon l'offre de services dans un environnement immédiat, quatre types de services ont été retenus : les supermarchés d'alimentation (sept grandes bannières; n = 167), les pharmacies de grande surface (n = 196), les bibliothèques (n = 55) et les banques (n = 4778, en fonction des huit bannières de banques et caisses différentes).

Le choix des services a été formulé selon divers critères. En comparaison aux plus petits commerces de produits alimentaires, les supermarchés contiennent une plus grande variété de produits et les prix sont plus compétitifs. Bien qu'ils ne représentent que 24% des commerces alimentaires sur l'Île de Montréal, ils détiennent 80% des ventes (Apparicio et al., 2007). Les supermarchés sont dispersés sur l'Île de Montréal, c.-à-d. qu'ils ne sont pas confinés au carrefour de routes principales et se retrouvent aussi dans des secteurs à plus forte densité de population et de plus grande mixité commerciale. Par exemple, la plus grande concentration de supermarchés sur l'Île de Montréal se retrouve dans le quartier du Plateau Mont-Royal (Apparicio et al., 2007).

Les grandes pharmacies ont été sélectionnées puisque la population à l'étude (adultes âgés de 45 ans et plus) est susceptible de fréquenter ces installations plus régulièrement, et qu'elles ont l'avantage d'offrir les services de pharmacies et une vaste sélection de produits à des prix compétitifs au regard des petites pharmacies. Les bibliothèques ont été retenues pour tenir compte des infrastructures culturelles des milieux, et du fait qu'elles sont fréquentées plus régulièrement par les populations locales que d'autres infrastructures culturelles comme les théâtres et musées. Par ailleurs, les bibliothèques se retrouvent souvent à proximité des maisons de la culture. Finalement, les banques et caisses populaires ont été retenues, car elles sont implantées dans tous les quartiers et sont donc plus dispersées dans la ville. De plus, mise à part le quartier des affaires, leur concentration est souvent en conjonction avec l'offre de divers services.

Les services ont été inventoriés et validés à l'aide de différentes sources documentaires, dont les sites web et les pages jaunes (Apparicio & Séguin, 2006). Les services ont été géocodés au rôle foncier c.-à-d. au centroïde du bâtiment, approche qui est plus précise que le géocodage selon le code postal. L'accessibilité aux services a été mesurée en calculant les distances aux services situés dans un rayon de 1000 mètres, et ce, sur le réseau routier (la distance réticulaire étant plus précise que la distance euclidienne) (DMTI Spatial, 2005). Les autoroutes n'ont pas été considérées dans les calculs d'accessibilité. Pour minimiser les erreurs d'agrégation (Apparicio et al., 2008; Hewko et al., 2002), les mesures d'accessibilité ont d'abord été calculées pour les centroïdes de chacun des îlots compris dans une

AD. Ensuite, la moyenne des mesures d'accessibilité pondérées par la population des îlots a été agrégée au niveau des AD.

Il est certain que l'accès à d'autres services aurait pu être examiné. Par exemple, il aurait été possible de considérer d'autres services tels les cafés et restaurants, les commerces de produits alimentaires spécialisés et de plus petites surfaces (boucheries, fromageries, fruiteries, etc.), l'accès aux stations de métro et arrêts de bus, ou encore la mixité de l'offre commerciale. Néanmoins, il est possible d'apprécier les services sélectionnés en tant que proxy de l'offre générale de services dans une unité spatiale donnée en raison de la co-localisation spatiale des différents services dans l'espace urbain.

La création de d'unités spatiales homogènes

Critères pour délimiter les zones

Trois critères ont guidé le choix de la méthode pour créer des zones homogènes. D'abord, les zones devaient être créées en fonction de la distribution spatiale des trois indicateurs du potentiel de vie active des milieux. La taille de la population de la superficie des zones ne fut pas considérée puisque les zones devaient être définies sur la base de la distribution spatiale de ces indicateurs. Ces indicateurs n'ont pas été regroupés dans un indice composite (tel que développés par certains, ex. Frank et al., 2005; Leslie et al., 2007) puisque les corrélations entre les indicateurs du potentiel de vie active étaient à la fois positives et négatives, ce qui aurait pu entrainer des erreurs de mesures du potentiel de vie active des milieux. Ensuite, la méthode pour créer des zones devait être optimale, c.-à-d. minimiser la variation intra-zone au niveau des indicateurs du potentiel de vie active et maximiser la variation inter-zones. Finalement, la méthode devait être rigoureuse, mais relativement simple à appliquer.

Délimitation d'unités spatiales : revue des méthodes possibles

Dans la littérature en géographie, plusieurs méthodes ont été développées pour créer des typologies d'environnement et délimiter des unités spatiales. Parmi celles-ci, notons les méthodes de tessellation, de « design automatique de zones» et la détection de clusters statistiques et spatiaux.

La tessellation réfère à la subdivision de l'espace en unités spatiales discrètes, c.-à-d. des unités contigües et qui ne se chevauchent pas. D'après Kemp (2008), il existe trois méthodes de tessellation : le carroyage en mailles régulières par lequel l'espace est défini en zones de taille et de forme identiques, les polygones de Thiessen / diagramme de Voronoï pour délimiter, à l'aide de la distance euclidienne, des zones d'influences autour d'un centroïde, et les « TIN » (triangulated irregular networks) pour modéliser l'environnement en trois dimensions.

La méthode de carroyage est certes intéressante et aurait pu être appliquée pour créer des zones homogènes. L'avantage de cette méthode est d'utiliser des unités de base sans *a priori* si ce n'est qu'au niveau de leur taille et de leur forme, par opposition aux AD qui ont un *a priori* au niveau de la taille de la population et de leur imbrication dans les niveaux statistiques supérieurs que sont les secteurs de recensement et subdivisions de recensement (http://www12.statcan.ca/francais/census06/reference/dictionary/geo021a.cfm). Néanmoins, le carroyage soulève quelques problèmes additionnels, notamment l'optimisation de la taille initiale des mailles (quelle superficie déterminer ?) et la possibilité d'introduire plus d'hétérogénéité dans la mesure des expositions.

Les AD ont été choisies comme unités fondamentales à la lumière d'autres études visant à délimiter des unités spatiales pour étudier les effets de milieux sur la santé. Ces études ont pour la plupart utilisé les plus petites unités statistiques disponibles, les « *enumeration districts* » au Royaume-Uni et les « *census block groups* » aux États-Unis. Par ailleurs, que l'on travaille avec la méthode de carroyage ou avec les

AD, les zones homogènes obtenues dépendront toujours des unités de base à partir desquelles elles auront été créées. Il serait certainement intéressant de comparer les deux méthodes. Mais à ce stade-ci de la thèse et à la lumière d'autres travaux visant le design de zones pour étudier les déterminants environnementaux de la santé, l'avantage de l'une ou l'autre des méthodes ne peut être déterminé.

Pour ce qui est de la création de nouvelles unités spatiales à l'aide des polygones de Thiessen, cette méthode permet, à partir d'un nuage de points, d'apparier toutes les localisations dans l'espace au plus proche voisin dans le nuage de points; le résultat est la génération de polygones convexes. Les polygones ainsi créés contiennent toutes les localisations les plus près d'un point d'origine, alors que les frontières représentent les localisations qui sont équidistantes entre les points d'origine. Cette méthode est souvent utilisée pour définir des aires d'influence (*catchment areas*) d'infrastructures ou de services.

Bien qu'intéressante, cette méthode ne permet pas de tenir compte de l'interaction de différentes variables spatiales dans le processus de création des zones. De plus, les frontières sont définies en fonction d'un point d'origine. Ainsi, pour étudier les effets de milieux sur la santé, la tessellation régulière est intéressante pour délimiter, par exemple, des zones en fonction du lieu de résidence des individus (une approche similaire est la délimitation de zones de proximité [buffers] en fonction du réseau de rues). Cette délimitation des milieux renvoie à la conceptualisation *idiosyncratique* de l'espace dans l'étude des effets de milieux, c.-à-d. que l'espace a une influence particulière pour chaque individu, et donc ne correspond pas à la conceptualisation de l'espace privilégiée dans cette thèse, soit l'espace intégré (le milieu est conceptualisé comme un niveau écologique d'influence dans lequel les individus sont imbriqués et exposés aux mêmes conditions). De plus, ces zones sont déterminées en fonction de la distance euclidienne. Il existe des méthodes plus précises de tessellation par lesquelles les zones sont définies à partir de la distance réticulaire (réseau de rues) et basées sur des fonctions globales d'analyse d'images, notamment les fonctions *CostAllocation et EucAllocation* (Cloutier, et al. 2007)

Le design automatique de zones réfère à l'emploi de logiciels avancés qui automatisent la création de zones en fonction de critères prédéfinis (ex : Cockings & Martin, 2005; Haynes, Daras, Reading, & Jones, 2008; Flowerdew, Manley, & Sabel, 2008). Or, bien que ces logiciels permettent de définir des unités spatiales dont l'homogénéité interne est optimisée, ils ne permettent pas une caractérisation multivariée des unités spatiales de base. Les variables en fonction desquelles les nouvelles unités spatiales seront déterminées doivent être combinées en un indice, ce qui est problématique dans le cas des indicateurs du potentiel de vie active.

Finalement, une autre approche est la détection de clusters statistiques et spatiaux. En géographie, la classification des environnements, aussi connue comme géodémographie (Harris, Sleight, & Webber, 2005), permet de combiner de l'information sur la localisation spatiale avec des données démographiques pour créer des groupes (clusters) ou typologies d'environnement. Les unités spatiales de bases sont regroupées à l'aide de méthodes de classifications statistiques, telles le K-means. Ces méthodes permettent de maximiser l'homogénéité interne des groupes et de maximiser les différences entre eux. Par contre, puisque la classification est statistique, les clusters créés ne sont pas géolocalisés. La détection de clusters spatiaux permettrait de localiser de tels groupes dans l'espace; toutefois, les logiciels actuels de détection de clusters spatiaux fonctionnent essentiellement avec des données ponctuelles et non des polygones (unités spatiales).

Puisque qu'aucune des méthodes mentionnées ci-dessus ne répondait aux critères pour définir des unités spatiales homogènes en termes du potentiel de vie active des milieux, une approche qui combine une méthode de classification statistique, le K-means, à la composante cartographique des SIG a été développée. Cette approche en trois étapes est définie en détails dans l'article 2 de la thèse, intitulé « *Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for designing optimal zones* ». Les étapes de la création des unités ne sont pas reprises ici.

Les AD ont été utilisées comme unités fondamentales à être agrégées pour délimiter des unités spatiales homogènes. Les AD sont de petites unités spatiales composées d'un ou plusieurs îlots (pâté de maisons dont les côtés sont délimités par des rues formant des intersections). L'Île de Montréal est divisée en 3222 AD dont la superficie moyenne est de 0.15 km^2 (min : 364 m^2; max : 18 km^2) et une taille de population moyenne de 562 résidents (min : 44; max : 2138). Dix-sept AD avaient une superficie inférieure à 500 m^2 (correspondant vraisemblablement une forte concentration de population (ex : des tours d'habitation); ces AD ont été fusionnées à l'AD dans laquelle elles étaient suitées.

Les AD ont été retenues comme unités fondamentales puisque ce sont les plus petites unités spatiales pour lesquelles les données du recensement canadien sont diffusées (Statistique Canada 2003). Pour des fins de comparaison avec les secteurs de recensement, les zones devaient être caractérisées en fonction du SSE. Au moment de la création des zones, les données populationnelles du recensement étaient seulement disponibles au niveau des AD, les micro-données du recensement n'étant pas encore disponibles.

Annexe III

Projet MARCHE : Sommaire exécutif

Résumé du projet

Environnements physiques et pratique de la marche chez les adultes

La sédentarité représente un enjeu de santé publique important car les coûts occasionnés par des maladies reliées à la sédentarité sont élevés, la prévalence de la sédentarité est importante et la prévalence de la sédentarité est plus élevée dans les populations défavorisées. Afin de développer des interventions populationnelles efficaces, il est donc primordial de mieux comprendre les déterminants individuels et écologiques de la pratique régulière de l'activité physique. En particulier, il existe d'importantes lacunes dans les connaissances en ce qui concerne l'association entre les environnements physiques du quartier de résidence et le niveau d'activité physique, et plus particulièrement, la marche. Une meilleure compréhension de l'association entre le potentiel piétonnier et la marche semble être une avenue prometteuse pour développer des interventions de type populationnel. L'objectif principal de ce projet de recherche était donc de développer un modèle conceptuel qui décrit l'association entre les environnements physiques et la marche dans des quartiers résidentiels situés dans des milieux urbains et de banlieue. À cette fin, nous avons construit deux bases de données qui ont ensuite été analysées par le biais de modèles statistiques novateurs. La première base de données a été constituée de données obtenus par observation sur les environnements physiques de 112 secteurs de recensement dans les 27 arrondissements définis selon la ville dans la région métropolitaine de Montréal. Des évaluateurs formés ont utilisé des grilles d'observation pour évaluer des aspects des installations physiques/environnementales de 112 quartiers (définis par secteurs de recensements) circonscrits dans les 27 arrondissements de la région métropolitaine de Montréal. Dans la deuxième base de données, près de 3000 personnes âgées de 45 ans et plus vivant dans les quartiers des arrondissements ont offert leur appréciation des différentes dimensions des installations physiques et environnementales de leur entourage et rapporteront leurs habitudes de marche dans le quartier. Les analyses statistiques ont : (i) documenter la validité et la fiabilité des mesures des installations physiques/environnementales, (ii) examiner les associations entre les installations physiques/environnementales et la pratique de la marche, (iii) examiner les associations entre les installations physiques/environnementales et différents indicateurs du statut pondéral; (iv) déterminer si les associations entre les installations physiques/environnementales et la pratique de la marche sont les mêmes chez les personnes ayant des incapacités physiques et celles n'ayant pas d'incapacités physiques. À partir de ces analyses collectives, nous avons développé un modèle des associations entre les environnements physiques du quartier de résidence et la marche dans des environnements urbains et de banlieue.

Annexe IV

Projet MARCHE : Approbation du Comité d'éthique sur la recherche chez les êtres humains de la Faculté de médecine de l'Université de Montréal

ARTICLE 2

ESTABLISHING THE SOUNDNESS OF ADMINISTRATIVE SPATIAL UNITS FOR OPERATIONALISING THE ACTIVE LIVING POTENTIAL OF RESIDENTIAL ENVIRONMENTS:

AN EXEMPLAR FOR DESIGNING OPTIMAL ZONES

Mylène Riva [1, 2, 3, 4]
Philippe Apparicio [4, 3]
Lise Gauvin [1,2, 3]
Jean-Marc Brodeur [1, 2, 6]

[1] Département de médecine sociale et préventive, Université de Montréal, Montréal, Canada

[2] GRIS, Groupe de recherche interdisciplinaire en santé, Université de Montréal, Montréal, Canada

[3] Centre de recherche Léa-Roback sur les inégalités sociales de santé de Montréal, Université de Montréal, Montréal, Canada

[4] Institut National de la Recherche Scientifique – Urbanisation, Culture et Société, Montréal, Canada.

[5] Analyse et évaluation des interventions en santé, AnÉIS, Université de Montréal, Montréal, Canada

Cet article a été accepté pour publication au International Journal of Health Geographics

Article 2. Rôle de chaque auteur

Riva, M., Apparicio, A., Gauvin, L., & Brodeur, J-M. Overcoming heterogeneity of exposures within administrative spatial units: an exemplar for designing optimal zones

Mylène Riva a développé les idées, la conceptualisation et l'opérationnalisation de l'approche pour créer des unités spatiales homogènes. Elle a fait les analyses, la cartographie des résultats et rédigé l'article.

Philippe Apparicio a collaboré à la conceptualisation et à l'opérationnalisation de l'approche, a supervisé l'analyse des résultats et a participé à la rédaction de l'article.

Lise Gauvin a collaboré à la conceptualisation de l'approche, a supervisé l'analyse des résultats et a participé à la rédaction de l'article.

Jean-Marc Brodeur a collaboré à la conceptualisation de l'approche et à la rédaction de l'article.

ARTICLE 3

STUDYING THE INFLUENCE OF AREA-LEVEL ACTIVE LIVING POTENTIAL ON WALKING:

CONTRIBUTION OF DESIGNING HOMOGENOUS SPATIAL UNITS

Mylène Riva [1, 2, 3]
Lise Gauvin [1, 2, 3]
Philippe Apparicio [2, 4]
Jean-Marc Brodeur [1, 2, 5]

[1] Département de médecine sociale et préventive, Université de Montréal, Canada

[2] GRIS, Groupe de recherche interdisciplinaire en santé, Université de Montréal, Canada

[3] Centre de recherche Léa-Roback sur les inégalités sociales de santé de Montréal, Canada

[4] Institut National de la Recherche Scientifique – Urbanisation, Culture, et Société (National Institute for Scientific Research – Urbanisation, Culture, and Society), Montreal, Canada.

[5] Analyse et évaluation des interventions en santé, AnÉIS, University té Montréal

Cet article a été soumis à Social Science and Medicine

Article 3. Rôle de chaque auteur

Riva, M., Gauvin, L., Apparicio, P., Brodeur, J-M. Studying the influence of area-level active living potential on walking: contribution of designing homogenous spatial units.

Mylène Riva a conceptualisé l'étude, fait les analyses et rédigé l'article.

Lise Gauvin a collaboré à la conceptualisation de l'étude et supervisé l'analyse des données. Elle a collaboré à la rédaction de l'article.

Philippe Apparicio a participé à la rédaction de l'article.

Jean-Marc Brodeur a participé à la rédaction de l'article.