

BMJ Open Neighbourhood environment factors and the occurrence of injuries in Canadian adolescents: a validation study and exploration of structural confounding

Afshin Vafaei, William Pickett, Beatriz E Alvarado

To cite: Vafaei A, Pickett W, Alvarado BE. Neighbourhood environment factors and the occurrence of injuries in Canadian adolescents: a validation study and exploration of structural confounding. *BMJ Open* 2014;4:e004919. doi:10.1136/bmjopen-2014-004919

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2014-004919>).

Received 23 January 2014

Revised 28 May 2014

Accepted 30 May 2014



CrossMark

Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada

Correspondence to

Dr Afshin Vafaei;
5av14@queensu.ca

ABSTRACT

Objectives: Social sorting mechanisms or analogous selection processes may impose confounding effects in the study of aetiological relationships. Such processes are referred to as structural confounding. If present, certain strata of social factors could hypothetically never be exposed to specific risk factors. This prohibits exchangeability across groups that is needed for meaningful causal inference. The objectives of this study were to: (1) develop and test the reliability and validity of composite scales for the measurement of social capital (SC), socioeconomic status (SES) and built environment (BE) and (2) to explore the possible roles of community level SC, SES and BE factors in studies of the aetiology of youth injury.

Setting/participants: A nationally representative sample of over 26 000 Canadian students aged 11–15 years.

Measures/analysis: Scales describing these key factors were developed and validated via exploratory and confirmatory factor analyses. We then used tabular analyses to explore structural confounding in our population.

Results: The proposed scales all demonstrated good psychometric properties. Despite variations in the number of adolescents across social and environmental strata, no evidence for the presence of structural confounding was detected in our data.

Conclusions: Relationships between social capital and the occurrence of injuries in Canadian youth aged 11–16 can potentially be studied without consideration of structural confounding biases. Canada is a suitable place to disentangle the effects of different neighbourhood social and environmental exposures on occurrence of injuries and other outcomes in adolescent populations. Exchangeability is possible across exposure strata and therefore a meaningful multilevel regression analysis is feasible. However, more studies are needed to test the consistency of our findings in other populations and for different outcomes.

Strengths and limitations of this study

- A national representative sample was utilised in this study.
- All contextual measures were re-validated to ensure their appropriateness.
- This is one of the few studies that explores the social epidemiology concept of structural confounding.
- Size of the communities and their rural/urban status were not considered in this study.

INTRODUCTION

An important potential bias in studying the impact of social factors on health outcomes in a multilevel setting is confounding resulting from social sorting mechanisms. This has been referred to as structural confounding.¹ When studying social factors at the community level, some strata of social variables may only contain participants who could never be exposed to aggregate level exposures of interest. For example, in a classic US example Messer² studied the effects of racial segregation on preterm birth and showed that few white women live in neighbourhoods with high levels of deprivation. In other words, some subgroups (white women) only experienced one level of exposure (low deprivation), and this referred to as ‘off-support’³ or ‘non-positivity’.^{4 5} Analyses of ‘off-support’ data in the presence of structural confounding rely on model extrapolation which do not allow examining the unique contribution of social factors, and thus limit meaningful causal inference.^{6 7}

There is no standard method for evaluating the presence of structural confounding. Tabulation analysis² is the method employed most commonly. In the presence of structural

confounding, disadvantaged people are sorted into communities with multiple disadvantages such as high neighbourhood density and poor street connectivity.⁸ As illustrated by studying the effects of neighbourhood deprivation on preterm birth in US populations,² when communities were stratified based on social factors indicating deprivation, percentages of populations by race and levels of education, there were very small numbers of deprived communities with high rates of university education or with low percentages of specific races. The presence of structural confounding can therefore be explored by stratifying communities according to the levels of social factors of interest and then examining the number of participants in combinations of the strata. In the case of existence of structural confounding, low numbers in some combination of social strata is not a function of sample size, it is due to the 'structure' of the society.

Community social capital is typically measured by levels of social cohesion and the quality of interpersonal relationships and has been shown to vary across other sociodemographic factors such as social class and education levels.^{9–10} Within child populations, independent associations have been reported between living in neighbourhoods with high levels of social capital and self-reports of improved general well-being,¹¹ emotional adjustment¹² and mental health.¹³ Evidence surrounding the impact of social capital on the occurrence of injuries in child populations is scarce. The few existing studies^{14–16} provide weak evidence on the preventive impact of social capital in occurrence of youth injuries. However, these studies all lack a well-developed and evidence-based conceptual framework and suffer from unrefined statistical analyses. None have explored the possibility of structural confounding effects of other community level factors.

Socioeconomic status (SES)¹⁵ and street connectivity,¹⁷ are each potential determinants of youth injury. Injury is a common and potentially preventable health outcome in child populations. Worldwide, more than 875 000 children die annually as a result of injury and non-fatal injuries are experienced by 10–30 million children and adolescents each year.¹⁸ Unintentional injuries are the leading cause of death for Canadian children and adolescents from 1 to 19 years of age.¹⁹ In order to estimate the unique contribution of community social capital on youth injuries, the effects of such community factors should be controlled for in multivariate analyses, and their influence as structural confounders requires consideration.

The aims of this study were (1) to describe selected individual and community level characteristics of Canadian school-aged adolescents; (2) to develop and test the reliability and validity of composite scales for the measurement of social capital, SES and built environment and (3) to determine the possible roles of community level SES, social capital factors, street connectivity and green space as structural confounding variables in such studies of the aetiology of youth injury.

METHODS

Data sources

Cycle 6 (2009–2010) of the Canadian HBSC is a cross-sectional survey of health behaviours and outcomes in a nationally representative sample of over 26 000 Canadian students aged 11–15 years. These students came from 436 schools in 11/13 provinces and territories in Canada.²⁰ The HBSC employs a multistage cluster sampling strategy in which the sampling units are classes nested within schools, which in turn are nested in their school boards, then provinces/territories. Institutionalised youth, youth in private schools, as well as those on the streets are excluded. Data collection involved administration of a written, in-class questionnaire that took 45–60 min to complete, administered from October 2009 until May 2010. The survey involves measurement of a variety of health topics including those pertinent to this study such as sociodemographic factors, self-reported occurrence of injury and indicators for interpersonal relationships.²⁰ The 2006 Canadian census data on indicators of SES such as family income, rates of employment and education were obtained from Statistics Canada.²¹ Built environment data were directly obtained via GIS technology.

Units of analysis

In this study, schools and their surrounding neighbourhoods were used as the aggregate units of analysis because we were interested in assessing the effects of built environment and area SES surrounding schools on the occurrence of injuries, as opposed to residential neighbourhood settings. Prior research also have considered the buffer around school reliable for constructs such as street types and connectivity, food environments, green space and socioeconomic environments.^{17 22–24} It has also been shown that in Canada, many of the characteristics of the school are important contextual factors associated with youth health behaviours and outcomes.²⁵

Measures

Individual-level characteristics

Injuries were estimated based on students' answers to a self-reported question: 'During the past 12 months, how many times were you injured and had to be treated by a doctor or nurse?' An ecological measure of injuries for each school was calculated by dividing the number of injured children over the total number of participants from the school.

Demographic variables of age, sex and individual SES were measured by direct questions. In concordance to similar studies,²⁶ family influence as a proxy for the individual level SES was measured based on the self-rated question about material wealth: *How well off do you think your family is?*

Community-level characteristics

Social capital

Children were asked to rate their feelings about five statements using response options from *strongly agree* to

strongly disagree. Statements were about 'trust' toward people around them, the possibility of asking for help or a favour from neighbours as a measure of 'cooperation', and three statements about 'cohesion'. Cohesion was defined as the level of interpersonal relationship and availability of safe places for social interactions and spending free time. Averages of individual scores were aggregated to the group (school) level as per existing precedents.^{27 28}

Socioeconomic status

We wanted to estimate the area SES around each school objectively and based on previous Canadian studies,²⁹ education, employment and average income are valid indicators for community SES. We used a 1 km buffer surrounding schools based on existing precedents.^{23 30 31} Previous research using the HBSC data has found this to be an appropriate buffer by which making meaningful inference about social constructs such as neighbourhood SES is possible.²³ Using census-based measures, education was defined according to the proportion of people (15+years) with at least a high school diploma and the employment ratio was defined as the proportion of people older than 25 who were employed. We also obtained the average family income for each buffer.

Built environment

Built environment indicators of street connectivity and green space were measured directly by employing GIS techniques for the same 1 km circular buffer, again as per existing precedents³¹; 1 km circular provides similar results (not identical) in measuring built environment factors when compared to road-network buffers²² and 5 km circular buffers¹⁷ and therefore is an adequate buffer for our purpose. Green space was defined as the proportion of land areas of parks, fields and wooden area relative to the total land area. Intersection density (number of intersections within the circular buffer divided by total area), average block length (sum length of roads within the buffer divided by number of real nodes), and connected node ratio (number of real nodes divided by all types of nodes such as intersections, dead-ends and cul-de-sacs) were used as indicators of street connectivity, as per existing precedents.¹⁷ Green space and street connectivity were viewed as separate built environment constructs.

Analyses

Descriptive

Key variables of interest were described at individual and community levels. Distributions of variables across different social capital groups were estimated and compared using analysis of variance and χ^2 tests.

Development of composite scales and testing their reliability and validity

Exploratory and confirmatory factor analyses were performed to assess the psychometric properties of

composite scales describing community social capital, SES and street connectivity. The robustness of the exploratory factor analysis was assessed by estimating diagnostic measure of sampling adequacy (the Kaiser-Meyer-Olkin measure) and by Bartlett's test of sphericity. Cronbach α levels for internal consistency were calculated as indicator of reliability.³² In order to provide a more rigorous assessment of the factorial structure and the validity of the derived scales, they were subjected to confirmatory factor analyses (CFA). Model fit was tested using the comparative fit index (CFI) and the Tucker-Lewis index (TLI)^{33 34} which findings of 0.95 or more were considered to indicate adequate fit.³⁵ The root mean square error of approximation (RMSEA) and standardised root mean square residual (SRMR) were used to estimate how well the model will reproduce the population covariance, with values less than 0.08 indicating a good fit.³⁴

The composite scale for social capital was defined as the sum scores of each item with the corresponding weights guided by loadings obtained from the factor analyses. To construct a composite scale for community SES, each of 419 schools was first ranked based on each of the three SES indicators. The schools were divided into tertiles and then were scored from 1 to 3 according to their tertiles. An additive composite scale for the SES was constructed including the scores of three indicators with the same weights. Similar methods were employed for constructing a composite scale for street connectivity. Since there are no meaningful cut-off points for our constructed composite scales of social capital, SES and street connectivity, we categorised them using quantiles for analytic purposes as per existing precedents.^{17 29} The ranges of composite scales were narrow (3–9 for SES and street connectivity; 5–25 for social capital) therefore we chose tertiles versus quartile or quintile²⁹ for categorisation.

Structural confounding

Stratified tabulations² were the main analytic tool for exploring structural confounding. Total number of students and number of occurrence of injuries in each combination of strata of community SES, built environment and social capital were estimated. According to the theories of structural confounding,^{1–5} non-existence or low numbers of students in *extreme cells* is suggestive of strong social stratification and possible structural confounding. We defined extreme cells as combinations of 'good' SES and 'good' social capital (highest tertile) but 'poor' built environment (lowest tertile), or combination of 'low' SES and low' social capital but 'good' built environment. There is no established *low number* to represent the existence of structural confounding. A standard low number might vary based on the context of the study and sample size. For example, Messer *et al*² in their study of neighbourhoods of two counties in North Carolina with a sample size of 31 715 considered fewer than 30 in each cell as evidence for structural

confounding. We did not set any number representing 'low number' a priori.

The Jonckheere-Terpstra test^{36 37} was used to establish if there were gradients in number of students by ordered groups of social factors. This is a sensitive non-parametric trend test for ordered differences among ordered groups, in which the null suggests that the distribution of the response variable does not differ across groups. All analyses were conducted using SAS V.9.3 (SAS Institute Inc, Cary, North Carolina, USA) and Mplus V.5.21.³⁸

RESULTS

Sample description

Table 1 demonstrates the distribution of individual and community level characteristics of the study population. From the total sample size of 26 078 students, 2546 individuals did not answer the social capital questions. Mean age (SD) of the participants was 14.2 (1.5), 48% were males, and 14.6% reported a non-sport-related injury in the last 12 months. As expected, injury rates in males were higher compared to females (15.4 vs 12.9, $p<0.0001$). Basic sociodemographic characteristics and

frequencies of injuries were not significantly different between those who responded to the social capital questions and those who did not (see online supplementary appendix table 1).

Census data for SES indicators at the 1 km buffer were available for 419 of the schools. The average family income for the entire sample was \$76 015 (SD=\$23 769), with suggestion of a normal distribution except for a number of outliers in high-income brackets. Mean proportions of people with more than high school education and with employment were 62% and 72%, respectively (table 1). Street connectivity indicators were available for most buffers, although green space data for 115 schools (27%) were unavailable due to lack of imaging. This occurred particularly in very remote areas of Canada such as in Nunavut and Northwest Territories where the main land area contains little vegetation and was mainly open space covered with clay, permafrost, rock, gravel and water and remote rural areas where there is mainly agricultural land.

Normal distributions observed for the SES and street connectivity indicators satisfied the main assumptions for factor analysis.

Table 1 Distribution of individual and community level characteristics in the total study population and by community social capital (SC) status

	Total	High SC (n=143)	Medium SC (n=140)	Low SC (n=136)	p Value*
Individual level characteristics					
Number of student	23 532	7799	8389	7344	
Age (mean, SD)	14.2 (1.5)	13.9 (1.5)	14.2 (1.5)	14.3 (1.5)	<0.0001
Per cent male	48.17	49.3	47.6	47.7	0.06
Number of injured (%)	3449 (14.7)	1142 (14.6)	1221 (14.6)	1086 (14.8)	0.92
Family affluence (%)					
					<0.0001
Well off	12 598(53.6)	3891 (49.9)	4544 (54.2)	4163 (56.7)	
Average	7978 (33.9)	2871 (36.8)	2828 (33.7)	2279 (31)	
Not well off	2223 (9.4)	827 (10.6)	778 (9.3)	613 (8.4)	
Missing	733 (3.1)	210 (2.7)	239 (2.8)	284 (3.9)	
Community level characteristics†					
Socioeconomic status indicators					
Average income \$ (mean, SD)	76 015 (23 768)	78 115(29 820)	76 553 (18 625)	73 215 (21 063)	0.21
Employed (mean proportion, SD)	0.62 (0.11)	0.61 (0.11)	0.63 (0.11)	0.63 (0.11)	0.19
High school diploma and higher (mean proportion, SD)	0.72 (0.13)	0.73 (0.13)	0.73 (0.12)	0.69 (0.14)	0.02
SES composite index	6 (2.1)	6.2 (2.1)	6.4 (2.1)	5.7 (1.9)	0.04‡
Built environment characteristics					
Intersection density (mean, SD)	40.6 (29.2)	40.6 (37.9)	38.4 (19.8)	42.9 (26.5)	0.43
Average block length (mean, SD)	283 (230)	308 (329)	278 (187)	262 (113)	0.23
Connected nodes ratio (mean, SD)	0.83 (0.10)	0.82 (0.10)	0.82 (0.11)	0.85 (0.09)	0.05
Street connectivity composite index (mean, SD)	6.0 (2.0)	5.9 (1.9)	5.8 (2.0)	6.3 (2.0)	0.07
Green space (mean, SD)	0.14 (0.14)	0.138 (0.14)	0.14 (0.14)	0.139 (0.13)	0.99
Population health status					
Average rates of injuries	0.14 (0.06)	0.14 (0.07)	0.14 (0.06)	0.13 (0.06)	0.39

*Comparing social capital groups by analysis of variance and χ^2 test when appropriate.

†At 1 km buffer of schools.

‡Low SC group significantly different from medium SC group (Bonferroni multiple comparison test).

Development and testing the reliability and the validity of composite scales

Exploratory factor analysis of social capital items indicated high and similar loadings for all included items, and a relatively high internal consistency (Cronbach's- α =0.76; figure 1). The statistical significance of Bartlett's test suggested that one factor was sufficient to explain the correlations. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was higher than 0.60 (0.79) which is tolerable and suggested that the dataset was appropriate for a factor analysis.³⁹ The confirmatory factor analysis indicated a good fit model, with a CFI of 0.968, a TLI of 0.974, a RMSEA of 0.084 (95% CI 0.080 to 0.089) and a SRMR of 0.026.^{33 34} Owing to relatively similar loadings, all items were given the same weight in constructing the composite scale.

The developed composite scales for SES and street connectivity were also reliable and valid. Both scales showed good internal consistency (Cronbach's- α of 0.74 and 0.70, respectively), acceptable KMO, and a significant Bartlett's test, which supported the appropriateness of the data and adequacy of the one factor solution (figure 2). Confirmatory factor analyses basically indicated a perfect fit (CFI=1, TLI=1, RMSEA=0.0, SRMR=0.0) for both indices.

Variations across levels of community social capital

Students living in schools with lower social capital were significantly older, more often female and less likely to live in families with average levels of affluence (table 1). Furthermore, schools with lower social capital were located in communities with significantly lower levels of education. Compared to areas with low social capital, average family income in high social capital areas was higher by 6.3% ($p=0.21$) and the composite index of community SES was significantly lower in schools with

low social capital compared to those with medium social capital (5.7 vs 6.4; $p=0.04$).

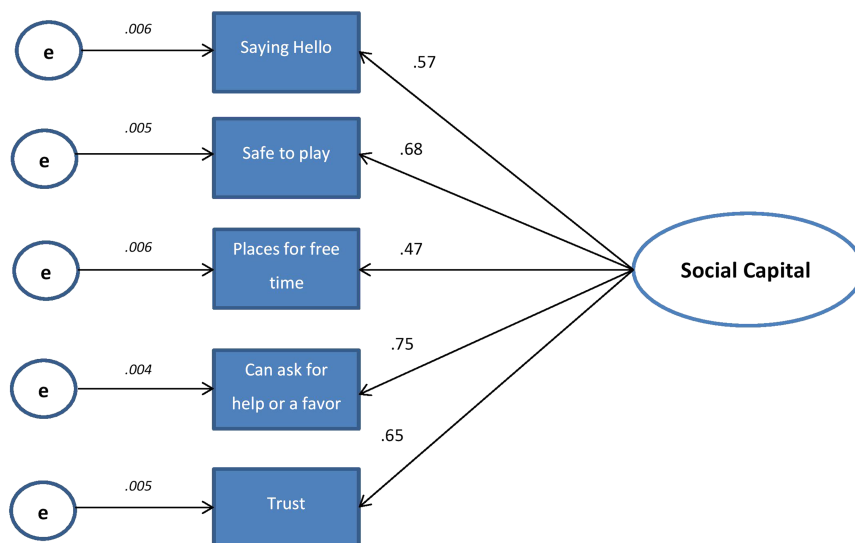
The differences between social capital levels for the proportion of area with green space was less than 1% ($p=0.99$) and for street connectivity was 7.9% ($p=0.07$).

Structural confounding

Tabulation analyses (tables 2 and 3) showed that there were no strong social sorting mechanisms at work for community levels of social capital, SES and built environment factors (either measured as street connectivity or as green space). We therefore observed no structural confounding effects. Student populations and injured cases were observed in all combination of strata of our three community measures and none of so-called *extreme cells* (defined above in methods) were empty or with sparse numbers of observation. This is indicative of potential absence of structural confounding and segregation in our population of Canadian youth. All cells contained the outcome and future multivariate regression analyses should be based on actual observed data not 'off-support' and 'smoothed over' cells.¹ We also performed the same tabulation analysis stratified by sexes. Unsurprisingly, the occurrence of injuries was more frequent among boys but applying the same methodology showed no evidence of structural confounding neither in boys nor in girls (data not shown).

The Jonckheere-Terpstra tests showed that significantly more students went to schools surrounded by poor street connectivity and poor green space. This was observed in all levels of community social capital, however was more pronounced in high social capital communities (tables 2 and 3).

Figure 1 Confirmatory factor analysis of the social capital measure.



Numbers at right represent factor loadings, italic numbers represent standard error of each item after fitting the CFA model, Cronbach's Alpha= 0.76, Overall KMO measure of sampling adequacy=0.79, P. for Bartlett's test <.0001, CFI=0.968, TLI=0.936, RMSEA=0.084 (0.080-0.089), SRMR=0.026

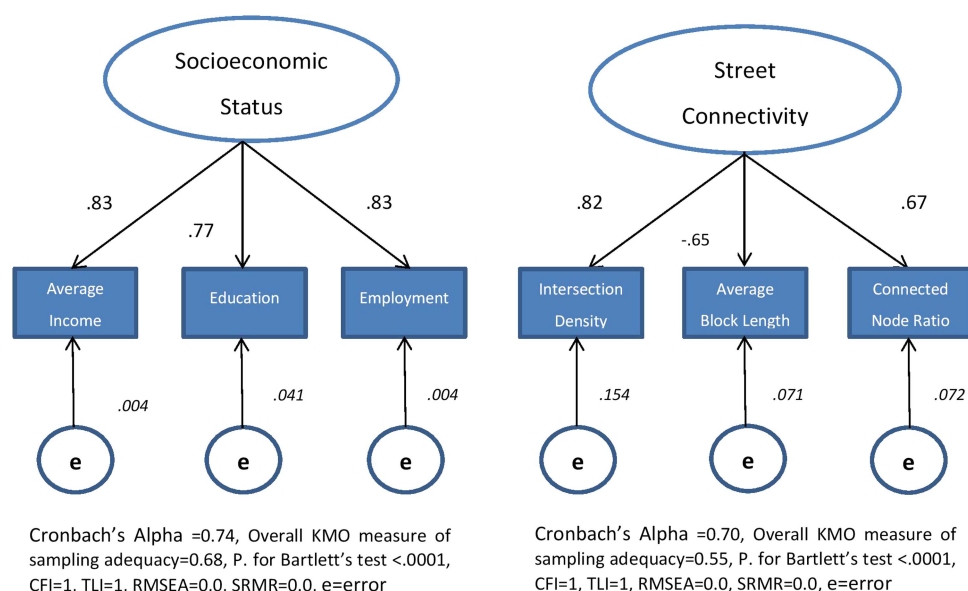


Figure 2 Confirmatory factor analysis of socioeconomic status and street connectivity measures.

DISCUSSION

In this study of structural confounding in adolescent Canadian populations, we chose to study three social and environmental factors of social capital, SES and built environment which have been shown to be potential risk factors for injuries in adolescents.^{15 17} All items used to develop composite scales for social capital, SES and street connectivity had been employed in similar studies previously^{8 15 17 31} but since validity is about the attributes of the people who are assessed not the inherent characteristics of the scale³² we revalidated all scales for our population. We employed three items of education, employment and income which were previously identified as the material deprivation indicators among Canadian population²⁹ as measures for community SES. Satisfactory results of exploratory and confirmatory factor analyses showed that these measures are reliable

and valid indicators for community SES in the study population. Age of 15 might not be the best cut-off point for measuring levels of education since people usually graduate from high school at the age of 18 or 19, however we were constrained to the use of aggregated census data provided by Statistics Canada. Nevertheless, our objective was to compare SES across schools and since the education measure was used uniformly across all schools the 'not optimal' age cut-off is not likely to produce any bias in our results. The street connectivity composite scale based on intersection density, average block length and connected node ratio in a 1 km radius around each school has been used in similar studies^{8 17} but was never validated. We therefore re-examined this scale in our analysis and it indicated a perfect fit.

We constructed the social capital composite scale based on the scale development methodology suggested

Table 2 Numbers of students and rates of injuries in each combination of social capital, SES and street connectivity

	High social capital			Medium social capital			Low social capital		
	Street connectivity			Street connectivity			Street connectivity		
SES	Good	Average	Poor	Good	Average	Poor	Good	Average	Poor
	Students	Students	Students	Students	Students	Students	Students	Students	Students
	Number of	Number of	Number of	Number of	Number of	Number of	Number of	Number of	Number of
	Injuries	Injuries	Injuries	Injuries	Injuries	Injuries	Injuries	Injuries	Injuries
High	487*	575	1082†	407†	845	846	509‡	477	816†
	87*	79	176	69‡	135	121	89	65	116
Medium	219	782	2317	1211	1248	1207	840	794	931
	38	112	347	176	198	192	143	128	138
Low	795†	571	810	970	572	911	498†	1227	1069
	132	101	129	164	89	135	85	236	136

*p of the Jonckheere-Terpstra test < 0.0001.

†Extreme cells (high SES and social capital but poor street connectivity or low SES and social capital but good street connectivity).

‡p of the Jonckheere-Terpstra test < 0.05.

SES, socioeconomic status.

Table 3 Numbers of students and injuries in each combination of social capital, SES and green space

	High social capital			Medium social capital			Low social capital		
	Green space			Green space			Green space		
SES	Good	Average	Poor	Good	Average	Poor	Good	Average	Poor
	Students	Students	Students	Students	Students	Students	Students	Students	Students
	Number of	Number of	Number of	Number of	Number of	Number of	Number of	Number of	Number of
	Injuries	Injuries	Injuries	Injuries	Injuries	Injuries	Injuries	Injuries	Injuries
High	459*	755	232†	697*	533	349	604*	330	348†
	67*	115	38	118*	78	49	85	43	62
Medium	850	1201	1002	1031	1369	899	842	269	826
	118	157	176	129	234	146	127	45	129
Low	358†	215	1334	538	609	1132	531†	1141	704
	45	43	232	86	87	182	101	189	102

*p of the Jonckheere-Terpstra test <0.0001.

†Extreme cells (high SES and social capital but poor green space or low SES and social capital but good green space).

‡p of the Jonckheere-Terpstra test <0.05.

SES, socioeconomic status.

by Streiner.³² First, in order to have appropriate content validity,²⁸ three main domains of social capital (cohesion, cooperation and trust) were measured by five questions in the HBSC survey. Then, by exploratory factor analysis we evaluated the factor structure of these domains and demonstrated that all domains loaded onto a single underlying factor. Cronbach- α of 0.76 for the five-item scale shows proper internal consistency (larger than 0.70)³² and also being smaller than 0.90 was indicative of no item redundancy.⁴⁰ High loadings of all of these five items showed that the items tap into the hypothetical construct of social capital³² and thus our scale has good construct validity which was further confirmed by acceptable model fit indices calculated by the confirmatory factor analysis.

In tabulation analyses, we found no evidence for the presence of structural confounding in the study population. All cells in our tabulation analyses contained data. This is a true indication of absence of structural confounding, not an artefact, due to our large sample size. A study performed by Messer² showed that in the socially stratified context of neighbourhoods of North Carolina, even with a large sample size of 31 715 there will be no people in extreme cells. She showed that literally no white woman lived in the most deprived areas and no black woman lived in privileged neighbourhoods. Our results showed that in Canada, there are fewer multiple social disadvantages and there are economically poor communities with good built environment or with high social capital. Statistically significant findings from the Jonckheere-Terpstra tests suggest that there is a trend in number of students according to levels SES and built environment factors. In all subgroups of social capital, it is more likely for students to go to school located in poor SES and unfavourable built environment communities, in other words, low SES and poor built environment communities are significantly more populous.

Our findings contrasted with those from similar studies which mostly have been performed in the USA.^{2 8} This absence of structural confounding in Canadian youth

populations has several potential explanations. Structural confounding from community social factors occurs more often in countries with higher levels of social stratification such as in the USA.¹ Compared to the USA, factors creating social segregation such as income inequality are weaker in Canada.^{41 42} We also did not stratify schools based on rural-urban status. Rural communities can enjoy high levels of social capital but at the same time suffer from poor built environment and low levels of SES. It is plausible that some of the schools with high social capital and low built environment are located in rural areas. Another reason for our null results may be due to our choice of aggregation. Schools may represent insufficiently refined geographical units to observe any meaningful variation across SES strata.

Limitations of this study warrant comment. Perfect fit of factor analysis models for SES and street connectivity can be of concern. It is possible that in confirmatory factor analysis models with small number of items lack of sufficient variations results in a saturated model with perfect fit³³; one interpretation is that the model is not reliable. We also made more parsimonious models (with only two items) and they also showed perfect fits. Our inability to fully consider the size of populations of communities is another limitation. Inequalities are at highest in large cities, separate analysis for large metropolitan areas might show different results and possible structural confounding within big cities. Our sample excluded certain groups of youth, mainly institutionalised individuals, those who live on the streets, and private schools. We were interested in studying community-related factors, and the exclusion of the first two groups who do not live in communities is not a problem for our study. However, exclusion of private school students is a potential threat to external generalisability of our results but since only 5% of students in Canada go to private schools, and less than 2% are in other types of situations (<http://www.statcan.gc.ca/pub/81-595-m/2013099/tbl1/tbl1.1-eng.htm>), this limitation is unlikely to produce considerable bias. Another threat to representativeness of our results is

the high percentage (27% of schools) of missing data for green space. These missing data were not random and were in very remote and rural areas where due to the characteristics of the environment (being mostly agricultural or non-vegetation) 'green space' is undefinable. Our results may not be generalisable to these areas.

In conclusion, in this study of school age Canadian students we addressed two major methodological issues of social epidemiological research: the validity of composite community level measures and the possibility of structural confounding. We showed our composite measures for social capital, SES and built environment are valid and also there is no evidence of structural confounding in our data. The aetiological analysis, that is, explaining factors contributing to variations in injury rates was not a purpose of this study. Owing to the non-existence of structural confounding in this population the independent effect of each community factor on occurrence of injuries, to be explored in subsequent multilevel multivariate analyses, can be estimated free of this major limitation of social epidemiological studies. However, more studies are needed to explore the existence of structural confounding in other populations and with different outcomes.

Acknowledgements The authors would like to thank the international databank manager Dr Oddrun Samdal, University of Bergen, Norway. The Canadian principal investigators of HBSC Dr John Freeman and Dr William Pickett, Queen's University, and its national coordinator is Matthew King. The authors also thank Mr Andrei Rosu for contributions in the GIS data collection.

Contributors AV developed the theoretical framework, performed the literature review, designed and conducted the statistical analyses. BEA and WP provided methodological and editorial supervision and guidance all through the study and contributed to the final manuscript.

Funding The Public Health Agency of Canada and Health Canada funded Cycle 6 of the Health Behaviour in School-Aged Children Survey in Canada. Additional support for this analysis included an operating grant from the Canadian Institutes of Health Research and the Heart and Stroke Foundation of Canada (MOP 97962; PCR 101415). International coordinator of the HBSC survey is Dr Candace Currie, University of St Andrews, Scotland.

Competing interests None.

Ethics approval The study protocol for the 2010 HBSC study has been approved by the General Research Ethics Board at Queen's University. Ethical approval for this particular analysis was obtained from the Queen's University Health Sciences and Affiliated Teaching Hospital Research Ethics Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Extra data including sex stratified tables of structural confounding is available by emailing Afshin Vafaei at 5av14@queensu.ca.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- Oakes JM. Propensity score matching for social epidemiology. In: Oakes JM, Kaufman JS, eds. *Methods in social epidemiology*. San Francisco: Wiley, 2006:370–92.
- Messer LC, Oakes JM, Mason S. Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. *Am J Epidemiol* 2010;171:664–73.
- Ahern J, Hubbard A, Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. *Am J Epidemiol* 2009;169:1140–7.
- Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc Sci Med* 2004;58:1953–60.
- Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 2004;58:1929–52.
- Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008;168:656–64.
- Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol* 2002;31:422–9.
- Oakes JM, Forsyth A, Schmitz KH. The effects of neighborhood density and street connectivity on walking behavior: the Twin Cities walking study. *Epidemiol Perspect Innov* 2007;4:16.
- Coulthard WA, Morgan A. *People's perceptions of their neighbourhood and community involvement. Results from the social capital module of the General Household Survey 2000*. London: The Stationary Office, 2001.
- Li Y. Social capital and social exclusion in England and Wales (1972–1999). *Br J Sociol* 2003;54:497–526.
- Runyan DK, Hunter WM, Socolar RR, et al. Children who prosper in unfavorable environments: the relationship to social capital. *Pediatrics* 1998;101(Pt 1):12–18.
- Stevenson HC. Raising safe villages: cultural-ecological factors that influence the emotional adjustment of adolescents. *J Black Psychol* 1998;24:44–59.
- Harpham T, Grant E, Rodriguez C. Mental health and social capital in Cali, Colombia. *Soc Sci Med* 2004;58:2267–77.
- Buckley LD. Social capital to prevent injury: providing skills for adolescents to protect their friends from risk-taking. *Proceedings of the Population Health Congress*; 6–9 July 2008, Brisbane, Australia.
- Elgar FJ, Trites SJ, Boyce W. Social capital reduces socio-economic differences in child health: evidence from the Canadian Health Behaviour in School-Aged Children Study. *Can J Public Health* 2010;101:S23–7.
- Huisman M, Oldehinkel AJ. Income inequality, social capital and self-inflicted injury and violence-related mortality. *J Epidemiol Community Health* 2009;63:31–7.
- Mecredy G, Janssen I, Pickett W. Neighbourhood street connectivity and injury in youth: a national study of built environments in Canada. *Inj Prev* 2012;18:81–7.
- Child and adolescent injury prevention: a global call to action*. Geneva: World Health Organization, 2005. http://whqlibdoc.who.int/publications/2005/9241593415_eng.pdf
- The Health of Canada's Young People: a mental health focus*. Public Health Agency of Canada, 2010. <http://www.phac-aspc.gc.ca/hp-ps/dca-dea/publications/hbcs-mental-mentale/assets/pdf/hbcs-mental-mentale-eng.pdf>
- The National Report. The Health of Canada's Young people: A mental health focus (PHAC, 2012).
- 2006 Census of Population. Statistics Canada. <http://www12.statcan.gc.ca/census-recensement/2006/index-eng.cfm> (accessed 15 Jun 2013).
- Seliske LM, Pickett W, Boyce WF, et al. Association between the food retail environment surrounding schools and overweight in Canadian youth. *Public Health Nutr* 2009;12:1384–91.
- Janssen I, Simpson K, Pickett W. Influence of individual- and area-level measures of socioeconomic status on obesity, unhealthy eating, and physical inactivity in Canadian adolescents. *Am J Clin Nutr* 2006;83:139–45.
- Huynh Q, Craig W, Janssen I, et al. Exposure to public natural space as a protective factor for emotional well-being among young people in Canada. *BMC Public Health* 2013;13:407.
- Davison CM, Russell K, Piedt S, et al. *the CIHR team in child and youth injury prevention among young Canadians: a national study of contextual determinants*. Vancouver, BC: CIHR team in Child and Youth Injury Prevention, 2013.
- Kukawadla A, Pickett W, Janssen I. Time since immigration and ethnicity as predictors of physical activity among Canadian youth: a Cross-sectional study. *PLoS ONE* 2013;9:e89509.
- Subramanian SV, Lochner KA, Kawachi I. Neighborhood differences in social capital: a compositional artifact or a contextual construct? *Health Place* 2003;9:33–44.
- Harpham T. The measurement of community social capital through surveys. In: Kawachi I, Subramanian SV, Kim D, eds. *Social capital and health*. New York: Springer, 2008:51–62.

29. Pampalon R, Raymond G. A deprivation index for health and welfare planning in Quebec. *Chronic Dis Can* 2000;21:104–13.
30. Nichol M, Janssen I, Pickett W. Associations between neighborhood safety, availability of recreational facilities, and adolescent physical activity among Canadian youth. *J Phys Activity Health* 2010;7:442–50.
31. Seliske L, Pickett W, Bates R, et al. Field validation of food service listings: a comparison of commercial and online geographic information system databases. *Int J Environ Res Public Health* 2012;9:2601–7.
32. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press, 2008.
33. Kline RB. *Principles and practice of structural equation modeling*. Guilford Press, 2011.
34. Schreiber JB, Nora A, Stage FK, et al. Reporting structural equation modeling and confirmatory factor analysis results: a review. *J Educ Res* 2006;99:323–37.
35. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equation Model-a Multidisciplinary J* 1999;6:1–55.
36. Jonckheere AR. A distribution-free k-sample test against ordered alternatives. *Biometrika* 1954;41:133–45.
37. Terpstra TJ. The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagat Math* 1952;14:327–33.
38. Muthén LK, Muthén BO. *Mplus user's guide: statistical analysis with latent variables*. Los Angeles, CA: Muthen & Muthen, 2010.
39. Cerny BA, Kaiser HF. Study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behav Res* 1977;12:43–7.
40. Boyle GJ. Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Pers Individual Differences* 1991;12:291–4.
41. Ross N. *What have we learned studying income inequality and population health?* Canadian Institute for Health Information, 2004.
42. Vafaei A, Rosenberg MW, Pickett W. Relationships between income inequality and health: a study on rural and urban regions of Canada. *Rural Remote Health* 2010;10:1430.