Author's response to reviews

Title: Associations between the built environment, total, recreational, and transit-related physical activity.

Authors:

Eric de Sa (eric_desa@hotmail.com)
Chris I Ardern (cardern@yorku.ca)

Version: 3 Date: 28 April 2014

Author's response to reviews: see over
Introductory Comment

The authors would like to thank the editor for the opportunity to respond to each reviewer’s comments.

We are pleased that Reviewer 1 agreed with our approach and has offered some helpful and straightforward suggestions. Reviewer 2 provided many insightful comments and raised a number of concerns to which we have subsequently responded. When appropriate, we have incorporated the recommendations into the paper or provided background rationale as to why including more information may deviate from the present scope. In particular, we would like to point out that since data is stored at the York University Research Data Centre (RDC), a data storage facility for Statistics Canada, there are strict policies surrounding the release of and re-analysis of data (especially with geospatial measures where residual disclosure is a concern).

Reviewer 1

Major compulsory

Comment 1: Please justify the use of a 500 m buffer zone around each centroid beyond the idea of approximate 5 minute walking distance. Wouldn't it have made more sense to use the whole postal code? If not, how do you think this would have changed the results if at all?

Response: Thank you, we have now clarified that the entire 6-digit postal code was included for these analyses (page 7, methods): “The geocoding process resulted in the formation of a centroid to represent each 6-digit postal code region.”

Both the first 3 digits belonging to the Forward Sortation Area (FSA) of the postal code, along with the final 3 digits, belonging to the Local Delivery Unit (LDU) were necessary to identify respondents of the CCHS belonging to the York Region Health Region. These analyses would not have been possible if only the FSA component was used since this would encompass a much broader population located in southern Ontario.

Minor essential

Comment 2: Page 5 in Methods. Why was 1% of random digit dialing used?

Response: The 1% random digit dialing (RDD) procedure is used as part of the overall sampling methodology which includes an area and telephone frame. Together, this sampling provides reliable estimates of more than 120 health regions, representing 98% of the population. RDD is used where the labour force survey sampling frames are much lower than required by the CCHS, where sampling costs are high, or where the quality of the list frame is considered poor (e.g. northern Quebec and Saskatchewan).

Comment 3: What does "some editing" mean?
**Response:** “Some editing” refers to the editing of data that is performed at the time of the interview by the computer-assisted interviewing (CAI) application. As noted on the CCHS website, “In response to some types of inconsistent or unusual reporting, warning messages are invoked but no corrective action is taken at the time of the interview.”

The following sentences were removed from the manuscript:

“Some editing of the data was performed at the time of the interview by the interviewer using the CAI application. It is not possible for interviewers to enter out of range values, and flow errors were controlled through programmed skip patterns.”

The following sentences were added to the manuscript:

“To reduce the number of errors in survey reporting, the CAI is not programmed to accept out-of-range values and flow errors are controlled through programmed skip patterns. For inconsistent or unusual reporting, warning messages are invoked and further edits are performed at the Head Office during the data processing step.”

**Comment 4: Were out of range values classified as missing?**

**Response:** When out-of-range values that are encountered during the CAI interview, the system gives a warning message and the interviewer asks the respondent to clarify their response to a value that is acceptable. If the respondent still gives a response that is unacceptable, then the value is inputted and subsequently flagged to be edited as either valid or missing by the head office.

**Comment 5: Page 9 in statistical analysis. Please describe respondents with missing data. What was the N? How much missing data was used as a criteria for exclusion?**

**Response:** The N value for the respondents missing data was quite low and could not be released by the Research Data Centre (RDC) since low frequency counts are handled cautiously due to the possibility of residual disclosure. The main criteria for respondents to be excluded from the analyses were instances when a main covariate (such as BMI or income) were missing and would therefore be automatically excluded from the regression models.

**Comment 6: Results-- how does the population here compare with the rest of the Canadian population for reference?**

**Response:** Because population weights were not applied to the analytical sample, it would be difficult to draw direct comparisons between these respondents from York Region and the rest of Canada.

Based on 2006 census data [http://www12.statcan.ca/census-recensement/2006/index-eng.cfm] which would be closest to the population in reference, this sample of York Region respondents surveyed by the CCHS was well-educated with 80% obtaining at
least some form of post-secondary education (versus 6 in 10 for all of Canada). Approximately 30% of the sample identified as being non-white compared to 16.2% of the Canadian population.

This region was selected with the aim to explore associations between physical activity and the built environment using a nationally representative survey specifically for a region outside of the commonly studied urban center. While the demographics of York Region are not directly representative of Canada, the variation of built environment that exists between different rural, suburban, and urban areas across Canada warrant the need to explore associations while adjusting for demographic covariates.

Comment 7: Page 11, is second quartile next to upper?

The upper quartile refers to the fourth quartile. The word “upper” has been replaced with “fourth” in the manuscript.

Comment 8: Would be helpful to have N in Table 1

Response: The analytical sample number has been added to Table 1.

Discretionary revisions

Comment 9: Does sample size at least warrant differences by gender?

Response: Gender, age and SES have all been found to have independent associations with physical activity. However, the purpose of these analyses were to show that even after adjusting for these factors, a relationship between physical activity and measures of the built environment still exists and can be identified using data collected by a nationally representative survey.

Future research may benefit from exploring various demographic differences with physical activity in the realm of built environment research. Gender differences in the association between perceived neighborhood safety and physical activity has been identified in previous work using pedometer-assessed activity data neighbourhood safety [e.g. Bennett et al. 2007]. In the future, it would be valuable to explore how different subpopulations (e.g. gender) experience the same built environment; however, focusing in on a specific area (or neighborhood) within York Region significantly reduces the sample size, and is beyond the scope of the current analysis.

Comment 10: Conclusion or discussion should mention safety. Do you posit that high density is potentially higher perception of safety?

Response: Studies investigating the influence of perceived neighborhood safety and PA participation are equivocal [Tucker-Seeley et al., 2009]. For example, one population-based study found no association between PA behaviours and perceived neighborhood safety for a group of adults in New South Wales Australia [Lim & Taylor, 2005], while
another study found that women who perceived their neighborhood as unsafe during night-time hours took fewer steps than those who viewed their surroundings as safe [Bennett et al., 2007]. Older adults who perceived their neighborhood environments as safe had higher leisure time PA levels than adults who did not share the same perception about their local environment and this association (between perceived safety and leisure time PA) persisted across socioeconomic strata [Tucker-Seeley et al., 2009].

By contrast, perceptions surrounding the built environment, as assessed by objective measures of safety (lighting near path and surveillance, etc.), are not consistently associated with walking in local neighborhoods [Pikora et al., 2006]. While it stands to reason that the method in which a built environment measure is captured (perception of measure versus objective method of quantification) impacts the strength of a relationship between the variable and different types of physical activities, in an area such as York Region where there are varying degrees of low- and high-density residential/commercial regions, the perceptions of safety associated with these areas could be highly variable. One limitation with exploring perceptions of safety in York Region is the notion that a suburban landscape is automatically more safe when compared to a more highly-dense metropolitan regions. While we agree that safety is an important contributor to human interactions with the built environment, these analyses are outside the scope of data that can be obtained by the CCHS.

The following sentence was added to address the issue of safety:

While perceived neighbourhood safety has been found to be associated with physical activity [Humpel et al., 2002; Pikora et al., 2006; Bennett et al., 2007; Tucker-Seeley et al., 2009], it could not be assessed within these analyses.

**Reviewer 2**

This well-written article employs a cross-sectional study design to assess the relationship between the built environment (BE) and two forms of physical activity (PA): leisure-time and transport-related PA. Participant data are abstracted from a large, national health survey and linked through geocoding to features of the built environment. Hierarchical linear and logistic regression modeling was employed to incorporate contextual neighbourhood features.

**Major Compulsory Revisions**

*Studies of factors associated with PA are important. However, there are a few major issues with this manuscript that jeopardize its potential to contribute to the field.*

**Comment 1:** The manuscript lacks novelty. There are now literally thousands of published studies focused on the role of the built environment and its contribution to PA. The majority of studies cited in the manuscript are, although important, quite dated (at least for this field) and there have been several major advances in theory and methodology since 2008-9. The results of the study are coherent with what is already known and are revealing of several possibly
problematic issues associated with the BE and health literature in general. For example, the claim that 3-28% of variation in PA participation is from neighbourhood differences in the built environment is somewhat misleading. Several recent reviews show that built environment effects are small or null and there is a large amount of variation among studies using similar methods. Moreover, these type of blanket statements require clarification about how neighbourhoods are defined; what about bias from non-publication of research with null or negative results? A study published in BMC Public Health by Oliver et al (2011, 11:959) is a good example.

Response: We thank the reviewer for this context and helpful suggestions. We do however, respectfully disagree with the assertion regarding the importance of this work, and feel that our analysis continues a critical public health discussion of the health-environment interaction related to physical activity. To this end, a simple PubMed search of ["built environment" AND ("physical activity" OR exercise)] yields 526 results, the overwhelming majority of which are not population-based nor devising a walkability index / methodological in nature. Within the Canadian context, there are no studies using CCHS in this manner, and limited research utilizing routinely collected secondary sources of data. This directly ties into the reviewer’s comment regarding “bias from non-publication of research with null or negative results?” – the current analyses are using a nationally representative survey that is widely used in other areas of health research for exploring associations between captured PA data with measures of the built environment. Should these results be considered “artifacts” (final comment in suggestion 2.c) as the reviewer suggests, then it would be helpful for future researchers to improve upon methodology with access to more data / employing new techniques. These analyses can serve as a reference for future research.

Comment 2: There are several conceptual and methodological issues that are either not incorporated or not addressed in the manuscript.

a) Use of the CCHS. Data from the CCHS are self-report and thus deserve close scrutiny, particularly data on health behavior and personal characteristics. Recent work on bias arising from self-report BMI data from the CCHS show that respondents are extremely likely to underestimate undesirable behaviours and characteristics (e.g. weight) and overestimate others (e.g. height) including time spent doing PA. While self-report is a characteristic of the dataset beyond the control of the authors, it certainly deserves mention in the manuscript either as a consideration in analytic procedures or as a limitation of the work. It was not clear from reading the manuscript that only data for York region were used in the analysis. This begs the question why the analysis was not performed on the whole data set, or at least for several population centres when data are available. Given the timing of the submission it would also seem prudent to combine data from the 2010-11 cycle to increase sample size. This should not be a problem given that each cycle is considered independent from each other. Moreover, the timing of the CCHS interviews is important in understanding PA. Participation in PA, particularly for leisure time and transport-related PA is dependent on season in Canada. When were the interviews conducted for this cycle of the CCHS or is the sampling distributed evenly temporally?
Response: The reviewer has made a number of important points which we agree will improve the transparency of the analyses and methods employed.

i) The rationale for studying York Region has been added to the manuscript, as follows:

“The Municipality of York Region is located directly north of Toronto and comprises nine municipalities: Town of Markham, City of Vaughan, Town of Richmond Hill, Town of Aurora, Town of Newmarket, Township of King, Town of Whitchurch-Stouffville, Town of East Gwillimbury, and Town of Georgina with the highest population growth rates occurring within the three municipalities closest to Toronto (Vaughan, Markham, and Richmond Hill). During the period of 1996-2001, York Region was the fastest growing Census Division in Canada with 30% of the population identifying themselves as visible minorities [York Region, 2003]. By 2010, the total population had exceeded one million people. From 1996 to 2001, there was a 30% increase in the employment labour force from 297,600 to 387,700. As of 2007, an estimated 485,000 people worked in York Region and is projected to increase to 800,000 jobs by 2031 [York Region, 2009]. With such a fast rate of growth, both in overall population and for jobs, York Region provides a unique opportunity to look at a demographically diverse population found outside of a major metropolitan area such as Toronto and to explore associations between the built environment and PA participation.”

ii) With regard to the issue of combining survey samples, the reviewer is correct in that the survey cycles are independent of each other. However, each survey cycle would need to be properly mapped to geospatial data for the given year. This is unfortunately a limitation of the resources available across York Region. Geospatial data is collected by each township or city before being sent to the central regional office which then compiles and edits it before academic institutions are able to request the data. As such, there is a lag time with the data available for analysis. Using the 2010-2011 cycle would not necessarily mean the most recent geospatial data were available and thus could skew any positive or negative associations with the built environment away from the true estimates.

iii) The reviewer poses the question as to why these analyses were not applied to all of Canada, or at least areas where data is available. Many studies utilizing surveys to ask questions on physical activity behaviours or using accelerometry measures are conducted in large metropolitan areas, often because of the readily available and reliable geospatial data. If individual municipalities or regions collect and release their geospatial data using different practices / release cycles, unnecessary variability is added to the analytical models that would not be adjusted for.

The paper cited by the reviewer (Oliver et al., 2011) in Comment 1 uses data collected from telephone surveys conducted in 2006 and census tract data from 2001:

- “The median family income of census tracts was obtained from the 2001 Census of Canada.”
- “The survey was conducted in February 2006, following a pilot survey in January 2006, and achieved a response rate of 29%.”
While technically the data used for our current analyses was collected between 2007-2008 and made available starting in 2010, we think the reviewer should be aware that better results can be obtained when the geospatial data and survey data are collected in similar years. It is possible that Oliver et al. ran into similar problems as did we in terms of matching up data to be representative of a closely-related population years.

iv) In regard to the issue of temporal sampling allocation, the following sentences will be added to the manuscript for clarification:

“The CCHS sampled respondents in all 9 municipalities of York Region during each month of the 2-year collection cycle. Although this mode of sampling is unlikely to impact on the spatial relationships observed here, future research would benefit by taking into account the month of data collection to account for seasonal effects on physical activity.”

Comment b): Geocoding. It is not clear why geocoding why PCCF+ software from Statistics Canada was not employed to assist with geocoding. The advantage of PCCF+ is its history in dealing with Canadian postal code data issues, such as: non-residential codes, population-weighted assignment (particularly important for more rural codes), imputations and the “vintage” codes problem. Larger postal code services areas exist because of smaller populations. Thus coding to a centroid will exacerbate misclassification of BE features to respondents. It would also be preferable to have more detail about the geocoding process. Using PCCF+ would also provide an opportunity to link respondents to socio-economic characteristics (potential moderators) of their neighbourhoods using small area-level census data. Inclusion of these data into the analysis would fit better with the idea of constructing a more “ecological” model of variations in PA and neighbourhood influences.

Response: It was suggested by the data librarian to use the same source for all geospatial data. Since the map information related to land-use and road-network came from DMTI, the postal code files would be most accurate to overlay. While coding to the centroid may be exacerbated using the current map data, it would nevertheless still exist if using the PCCF+ files since the specific location (number and street side) was not available using CCHS data, meaning that an approximation would still need to be determined based on the mean point between all areas of the postal code region.

Comment c): Definition of neighbourhood. Many studies of PA now use global positioning system data loggers and accelerometers to assist with identification of locations where objectively-measured PA occurs. These studies combined momentary (experience sampling) assessments are more powerful in their ability to investigate aspects of the built environment associated with PA. The methods used, in this study, to attribute respondents to features of the built environment are rudimentary and require attention. First, the true location of any respondent is dependent on the location of the postal code centroid in relation to their home address. At best this would be a city block. While the rationale for using a 500 m buffer is logical, it would seem prudent to take into consideration recommendations from several methodological reviews as part of the analytic strategy employed in the manuscript. For
example, walking distances will vary with the morphology of the land, with age, sex and other demographic characteristics. The authors should consider the creation of multiple buffer sizes to evaluate the influence of buffer size in the analysis. Furthermore, the 500 m rationale does not work if the trip (walking or cycling is one-way or if there is a substantive amount of time passing between the initial trip and the return trip). Second, buffers should ideally be created from street network distances rather than Euclidean or “as the crow flies” distances. Respondents who are walking or adopting active transportation will be using the street network (mainly). It is conceivable that the statistically significant results derived from analysis may be an artifact or an error in the process by which both PA and characteristics of the built environment are measured and attributed to CCHS respondents.

Response: The reviewer raises several important issues regarding the definition of “neighbourhood” that are the focus of ongoing investigation. While the reviewer has noted several complementary approaches, these approaches have their own limitations. One advantage to the use of CCHS data is that it is a routinely collected population-based survey capturing a wide variety of health outcomes and behaviour. However, we agree that the ongoing development (and validation) of more precise analytical techniques to (objectively) capture features of the built environment and physical activity offer an exciting opportunity for future work.

The authors also agree that using street network distances (rather than Euclidean distance) are the preferable means for determining buffer zones. However, this technique works best when the precise address (postal code, street side, residential number) is readily available and can be mapped. The Master Data File of the CCHS only contains the postal code. As such, the Euclidean distance should be interpreted as a measure the local neighborhood environment and not the maximum distance that can be travelled.

Comment d): Addressing potential for self-selection bias (Handy et al., 2006). Cross-sectional studies are particularly vulnerable to residential self-selection bias resulting from unmeasured neighborhood selection factors related to built environment exposures and PA. Neighborhood selection factors may include preference for PA resources, which could affect neighborhood choice and PA level. Similarly, social and financial resources not only influence where individuals can afford to live but also shape perceived barriers to PA. Furthermore, traditional covariate adjustment cannot adequately control for neighborhood preferences and other residential selection factors that are difficult or impossible to measure.

Response: We thank the reviewer for highlighting this important issue. We agree that self-selection is a critical factor, as is safety, access to amenities, and general aesthetics that tend to cluster in areas of high socioeconomic status. This issue is further complicated by the fact that lower income areas have the same number of “fee for use” recreational services, but fewer parks and free use features in the proximal built environment, a rallying point of the environmental justice movement. Work from Frank et al. (Soc Sci Med, 2007), however, shows variability in walking behavior and obesity rates amongst those who are discordant for walking preference (vs objectively assessed neighbourhood walkability), which highlights the influence of the built environment beyond individual preferences that may have important population-health implications.
We have added the following sentences in the manuscript to address this issue:

"In addition, as with all survey data, social desirability bias should be acknowledged. Finally, it is important to note that respondents are not asked the reason for selecting their current address as the question is not addressed during the CCHS interview process. It is unclear if the resident has selected their household for reasons based on neighborhood safety, proximity to schools, traffic flow, socio-cultural factors, aesthetics, or a variety/combination of many other factors that relate to homeownership and neighborhood dwelling”.

Comment e): Selection of built environment features. It is unclear given the PA measures of interest why walkability was not selected as an important feature of the built environment. Other BE variables that could be derived easily include cycling infrastructure (bike lanes, etc.) as well as public transportation routes (bus routes, stops, etc.) all of which have been shown to be related to walking and cycling-related PA.

Response: While we agree with the reviewer that the exploration of additional built environment features would have been valuable, not all of these indicators were available for York Region, and the data that was available did not correspond with the given CCHS data cycle. However, because there are no bylaws requiring use of bike lanes, and relatively few dedicated bicycle lanes during the time of analyses, in York Region it would not have been possible to obtain reliable geospatial data related to road networks with cycling infrastructure included.

Minor Essential Revisions

Comment 3 Page 5 (Exclusion criteria): It is not explained why only respondents from York Region were included in the analysis. What is the rationale for this decision and how would this criteria affect the results of the research?

Response: Please see Comment 2.a) for the additional paragraph related to why York Region was included in the analysis.

Comment 4. Page 7, Par 1 and 2: The second and third paragraph repeats text from the first paragraph and should be deleted.

Response: Thank you. Changes made.

Comment 5. Page 13: It is surprising that no mention is made of more recent state-of-the-art studies of BE and PA that collect objectively-measured PA and that have accurate delineations of neighbourhood boundaries. While there are some strengths to using the CCHS, these strengths can easily be outweighed by the several disadvantages to using CCHS data, e.g. respondent bias, self-report data, lack of accurate addressing and other issues.

Response: Please see other responses regarding study limitations and context. While the limitations of self-reported physical activity are well known, differential bias has not been consistently observed, and objectively assessed activity remains an emerging
science that is subject to its own limitations (e.g. cut-point bias, wear time, incomplete activity capture, etc.).