PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The association between the neighborhood social environment and obesity in Brazil: a cross-sectional analysis of the ELSA-Brasil study
AUTHORS	Chaparro, M. Pia; Pina, Maria; Cardoso, Letícia; Santos, Simone; Barreto, Sandhi; Giatti, Luana; Matos, Sheila Maria; Mendes da Fonseca, Maria; Chor, Dora; Griep, Rosane Haerter

VERSION 1 – REVIEW

REVIEWER	Elena Carrillo Álvarez Blanquerna School of Health Sciences, University Ramon Llull. Barcelona, Spain.
REVIEW RETURNED	28-Oct-2018

GENERAL COMMENTS	I commend the authors for a well-written and clear paper on the influence of the social environment on obesity in Brazil. I would like to encourage the authors to undertake a more thorough discussion of the findings in terms of the mechanisms through which the investigated association is different by gender and SES. Similarly, I would like to see the results discussed in terms of implications for male population. The authors may want to refer to Carrillo-Alvarez et al. (2018) for a conceptualized vision of how the social environment may influence obesity. https://onlinelibrary.wiley.com/doi/10.1111/obr.12760

REVIEWER	Lisa Kakinami and Anna Smyrnova (co-review) Assistant Professor, Concordia University, Canada (LK) Graduate student, Concordia University, Canada (AS)
REVIEW RETURNED	05-Nov-2018

GENERAL COMMENTS	Dear editors of BMJ Open, Thank you for the opportunity to review Chaparro et al.,'s manuscript entitled, "The association between the neighborhood social environment and obesity in Brazil varies by gender and neighborhood socioeconomic status: a multilevel study". The authors present a well-written investigation into how neighbourhood environments may affect obesity risk. Nevertheless, we have a number of comments and suggestions to the authors for their consideration. In general, key details on many methodological aspects are currently missing from the manuscript, interpretations of classifications (which are based on tertiles) are necessary to give readers a better sense of the magnitude of risks and associations,

	and a more thoughtful discussion/conclusion on what tangibly can be gleaned from this study (rather than a recommendation to 'reduce the perception of violence' which is a bit simplistic and insincere) is needed. More specific comments are outlined by section below. Abstract: Some methodological details are lacking. In addition, we feel the conclusion goes beyond what can be concluded from this specific study. In particular, we would caution against using the term "prevent", as this study reports the association in cross-sectional analysis, and not longitudinal or interventional research. Introduction: -The connection between obesity and PA, as well as obesity and socioeconomic status is not well established in the introduction. Further details of these are needed to properly 'set-up' the research study. (For instance, parts of the information on page 17, line 23 could instead be moved to the introduction). -Page 5, Line 22: The introduction alludes to the potential importance of interventions in vulnerable populations. However, as this study is based on civil-servants (which excludes the extremely poor and unemployed) it is not clear how the sample can be considered as a vulnerable population. Some explanation, or rephrasing of that sentence could help. -Page 5, line 29: Some statistics on the recent increase in obesity observed in Brazil (in %) would be helpful here. It would better illustrate why the research question is focused on obesity weight status, rather than on overweight/obesity. Methods: -It would be helpful to provide some statistics on how the current sample differs from the general population (distribution of income, skin color, education attainment etc) to help the readers better understand the generalizablity of study results. Similarly, statistics of the six different cities in this study would be helpful. These tables could be presented in the Appendix. -The authors develop their SES measure based on methodology described by Santos et al. (2010). However, Santos et al., defined SES based on four census tract indicators: income, education, persons per household and % in the 0-4 year age bracket. In contrast, this study defined SES not with education, but with percent of white residents. This change in methodology needs to be rationalized and explained. Related to this point, it is always preferable to provide the readers with enough relevant methodological details to understand the study, rather than referring them to an outside publication (such as the Santos et al., study). -As neighbourhood SES indicators were used for the analysis, more information on the distribution of participants within neighbourhoods is warranted. For instance, it is not clear whether authors used circular or network buffers centered on the participants' locations (to

	ensure that participants were not between two different neighbourhoods). If not, more details on whether participants were on the 'edges' of two adjacent (but very different SES) neighbourhoods is needed.-The test-retest values themselves should be provided, rather than a general statement that there was 'good internal consistency'. -The paragraph explaining the quality of scales (Page 8, bottom paragraph) could be moved to right after the sentence when they are first introduced (Page 7, line 45). -As the authors found significant sex differences, were the individual-level scores on social cohesion, perceived safety, and perceived violence also created separately for males and females? If not, rationale in why this was aggregated without adjustment for gender is warranted. A brief description and interpretation of the questionnaires' tertiles would be very helpful. For example, what an average value of 10.0 on a social cohesion scale for those in the highest tertile means in terms of social cohesion. (and similarly for perceived safety, and perceived violence) -Similar to the above point, it would be helpful to provide more description of the three SES categories, and how they differ from each other. (A brief table in the Appendix will help the interpretation of the results). For example, what does, "women residing in high and low SES neighbourhoods" mean in terms of the characteristics that these neighbourhoods have? What is the range? In particular, as the study has an inherent limitation of likely excluding the extremely poor population, a range of SES categories provides this important information. -It is not clear why the educational attainment categories of "less than primary" and "primary" were both retained in the regressions. Is this a meaningful comparison group to describe? Were there in fact significant differences noted between these two groups? Otherwise, the 'non-linear trends' shown between these different educational categories in the results (eg: table 2, model 3, greater risk was found for those with a secondary education, than with a primary education) needs to be mentioned in the discussion. -Page 10, line 17: "neighborhood level scores of our three neighborhood variables were reconverted into tertiles within each neighborhood SES category" – were any significant differences in scores noticed? Could you provide any figures/table about the distribution of SES categories and distribution of scores within each SES category? (it's interesting to see if, for example, score for violence was 'XX times less' in high SES than in low SES) Results: -It's not clear from Table 1 whether the distribution of age, education and skin color (etc) differ by sex. As the paper makes the argument for important sex differences, further information on these sex differences should be assessed and provided for the readers.

	-Were there any correlations between the various neighbourhood scores with one another? Is there any concern that they may be multicollinear with one another and so are presenting 'redundant' information in some of the regression models? This may have implications on including them in the same (multilevel) model, such as presented in table 5. -The article cited by authors (Sampson et al. 1997) suggested that "a measure of collective efficacy (defined as social cohesion among neighbors…) … is negatively associated with variations in violence" – was this association checked in this study? Discussion: -As mentioned previously, some explanation of the different SES tertiles will help with interpreting what 'high' and 'low' SES neighbourhoods mean (eg: Page 17, line 20) -The authors mention that gender related differences in neighbourhood perceptions have been previously reported in the literature. Did the authors also find this to be the case in this study? -Even though perception of violence and 'actual' level of violence (based on crime statistics, for example) are likely associated, we'd caution the authors against using these terms in an exchangeable way, especially since objective crime measures were not incorporated. Some neighbourhood interventions could be proposed to reduce the actual violence, but violence perception depends also on many other personal factors. As the authors cannot obtain objective measures of crime, references linking the association between actual violence and perceived violence would be warranted here (Page 20, first paragraph in particular). -The conclusions could use a bit more rumination. We feel it is an overgeneralization to conclude that increasing social cohesion and violence perception could be protective against obesity. This type of conclusion is a bit simplistic, and may be misconstrued as flippant and insincere. A more thoughtful conclusion would incorporate more tangible suggestions on what policy, researchers, or communities can do to try to help combat this problem together. Tables: -Table 1: The percentages should be the 'column' percentages, and not the 'row percentages'. IE, it should be the percent of women who were 'not obese' compared to the men who were 'not obese', rather than the percent of all women who were 'obese' compared with the percent of all women who were 'not obese'. Statistical comparisons should be done, and p-values should be in the table. Minor corrections: Abstract, Line 43: High SES and low SES neighbourhoods are mentioned, yet the use of tertiles is not described. The total range is

	not clear. Possible suggestion is to use “highest SES out of three tertiles” or “high and low tertiles”. Page 7, line 38: typo – (BMI ≥ 30) Table 2: Typo in the title (BMI ≥ 30) Table 3: Typo in the title (BMI ≥ 30) Table 4: Typo in the title (BMI ≥ 30) Table 5: Typo in the title (BMI ≥ 30)

REVIEWER	Ruben Brondeel Ghent University - Department of Movement and Sports Sciences
REVIEW RETURNED	28-Jan-2019

GENERAL COMMENTS	This article investigates important relations between neighborhood characteristics and BMI, in the context of social inequalities (gender and SES). The results are based on a very interesting data set, in terms of its origin, of its size and – despite the inclusion of only civil servants – of its composition. However, I have some comments on the methods. The analysis can be improved, which would make the results more trustworthy. Major comments: - The authors took decisions in the model building that are not motivated in the article and, in my opinion, not appropriate. o The dependent variable, BMI, was dichotomized. This decision is not motivated, and leads to a loss of information and power. Dawson, Neal V., and Robert Weiss. "Dichotomizing continuous variables in statistical analysis: a practice to avoid." (2012). o The independent variables were categorized in tertiles. Similar to dichotomizing the dependent variable, this has considerable consequences. Mabikwa, Onkabetse V., et al. "Assessing the reporting of categorised quantitative variables in observational epidemiological studies." BMC health services research 17.1 (2017): 201. Collins, Gary S., et al. "Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model." Statistics in medicine 35.23 (2016): 4124-4135. Kahan, Brennan C., et al. "A comparison of methods to adjust for continuous covariates in the analysis of randomised trials." BMC medical research methodology 16.1 (2016): 42. o The model presented in Table 5 includes the 3 neighborhood perception variables, controlling the estimated associations of each perception variable for the other two perception variables. In Table 2 to Table 4, the associations of the 3 variables are tested separately. The two different approaches are not motivated. I suggest to use the modeling strategy from the model in Table 5 throughout the article. o Moderator effects are tested in sub-group analyses. Introducing interaction terms in the regression model would result in less estimates and therefore in more precision in the estimations. o There is no motivations for the model building in Tables 2 to 4 (i.e. adding variables in the different models). This type of model building could be used as a rudimentary way of understanding mediation mechanisms. However, there are no such objectives in the article. And if there were, the sequence of introducing the variables would be different, with the perceptions being introduced in the final model.

	- The variables ‘perceived safety’ and ‘perceived violence’ are highly related, not only by concept but also by construct. One of the three indicators of ‘perceived safety’ is ‘Violence is a problem in my neighborhood’, which could be clearly an indicator ‘perceived violence’. The authors should give more information (e.g. correlation coefficients between 3 perception variables) and explanation (e.g. origin of scales) to motivate the distinction between the two constructs. Alternatively, a variable can be removed from the analysis. Minor comments: Introduction: - Social environment is loosely defined. What other aspects are often included in social environment, next to those included in this article? And how was the choice for ‘social cohesion’, ‘perceived safety’ and ‘perceived violence’ made? - Why is the adjective ‘perceived’ not used for ‘social cohesion’? I do not see a difference with ‘safety’ or ‘violence’. Methods: (note that some of the comments below might become irrelevant if the major comments are addressed) - (p7 - line 26) The mean can be easily calculated from the data in the line above. It would be more informative to report percentiles (e.g. the median, quartiles, minimum and maximum, …), especially because a skewed distribution can be expected. - (p9 – line 3) The aggregate of the perceived measures were used as neighborhood characteristics. I suggest to investigate both the individual and aggregated variables in the same regression model, to test the neighborhood effects above and beyond the individual effects. - Since there is a strong theoretical difference in perceived safety/violence between women and men, should you use gender-specific aggregated measures? - What are the Intra-class-correlation coefficients for the three perception variables? You could estimate these in multilevel models with each of the variables (before aggregation) as a dependent variable, a random effect for neighborhood, and no independent variables. Even though ICC is not a perfect measure, it would be an indication of how appropriate the aggregated measures are. - (p9 – line 29). The construction of neighborhood SES is not motivated. Why was there no cluster analysis directly on the 4 indicators? Why was there a cluster analysis at all? With the high number of participants included in the study, there is no need for variable reduction, and the 4 SES variables could have been used separately. Using the 4 variables in their original form (in combination with interaction terms instead of the sub-group analyses), would be more informative and would make it easier to reproduce the analyses. The sequence of variable reduction methods make the resulting SES variable very specific to the dataset at hand. - (p10 – line 5). Are the ELSA sites the 6 cities? If so, you could consider introducing the variable as a fixed effect instead of a random effect. This would provide more information and simplify the analyses. Results: - p10 – line 30: The second ‘sample’ should be deleted.

	- p10 – line 33: Use explicitly the age boundaries for 'middle aged people', because the term 'middle-aged' has many different uses. - P11 – line 39: The result presented here indicates the lack of a moderation effect (same effect, independent of the SES status). An overall dose-response association (i.e. a linear association) could be derived from Table 4, or a group-specific dose-response association could be derived in Table 5 for each group specific. - Table 5: What are the sample sizes in the 6 different samples? For women in low SES neighborhoods, there are some effect sizes which are not statistically significant, but they are as high as some estimates that were statistically significant in the other tables. The reader cannot derive if this is due to a lack of power related to the sample size, or to a higher standard deviation. Discussion and conclusion: - In the strobe statement, you mention that the study is explorative. However, the discussion section (nor the methods) does not reflect this. For example, p18 – line 11, the results from two out of three variables (and only one in the subgroup analyses in Table 5) are generalized to 'the neighborhood social environment'. I suggest that the discussion stays a little closer to the results, and indicates what is missing in the results (e.g. other dimensions of social environment, objective measures of social environment). STROBE statement: - Title: the term 'multilevel' rather reflects the type of analysis than the study design. 'Cross-sectional study' or 'cohort study' would be more in its place. - Descriptive data: the comment that complete data is used, is not an alternative to describing the (relatively high number of) participants with missing data. The reader needs to understand why some people had missing data and were therefore left out the analysis.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:
Reviewer: 1
Reviewer Name: Elena Carrillo Álvarez
Blanquerna School of Health Sciences, University Ramon Llull. Barcelona, Spain.
1. I commend the authors for a well-written and clear paper on the influence of the social environment on obesity in Brazil.

Thank you for this comment, we are pleased to read that you find our paper interesting and well-written.

2. I would like to encourage the authors to undertake a more thorough discussion of the findings in terms of the mechanisms through which the investigated association is different by gender and SES. Similarly, I would like to see the results discussed in terms of implications for male population.

We believe we have extensively discussed the possible differences by gender in the Discussion. However, based on the reviewer's comment we have expanded our discussion on gender differences and included a more in-depth discussion of possible mechanisms by SES.

As for the implications for men, we did not find any significant associations between the neighborhood social environment and obesity among men. Therefore, our conclusions related to the possible beneficial effects on obesity of interventions targeting the social neighborhood environment only apply to women.

Our findings are in line with other research using data from ELSA-Brasil. For example, a manuscript looking at the association between the food and physical activity environments and obesity with ELSA-Brasil participants found significant associations for women but not men.[1] Another study in the south of Brazil also found no effect of area-level education on BMI, waist circumference and obesity for men, whereas a significant effect was found for women.[2] Further, in our sample, demographic characteristics that are usual predictors of obesity also were not associated with obesity among men; namely, age and education. This is also in line with what is found for the Brazilian population at large. A recent report from a representative sample of the adult Brazilian population showed that while schooling was inversely associated with obesity among women in Brazil, there was no association between schooling and obesity among men.[3] The same study did find variations in obesity prevalence by age group for both men and women.

A paragraph related to the null findings among men has been added to the Discussion as follows: "A previous study conducted in the south of Brazil found neighborhood-level variations in obesity prevalence for both men and women; however, neighborhood-level education was only associated with obesity among women in the sample.[37] Another study using ELSA-Brasil data found that the food and physical activity neighborhood environments were associated with obesity among women but not men.[38] The results of these studies and our own suggest that the neighborhood environment may matter for men's obesity risk, but the neighborhood factors studied to date are relevant only for women. Future studies should further investigate which neighborhood factors, if any, affect obesity risk among men in Brazil and other Latin American settings, as well as the reason why neighborhood factors may affect women's and men's obesity risk differently."

3. The authors may want to refer to Carrillo-Alvarez et al. (2018) for a conceptualized vision of how the social environment may influence obesity. https://onlinelibrary.wiley.com/doi/10.1111/obr.12760

Thank you for providing us with this reference. It has been incorporated into the Introduction as an additional reference for the statement "The social environment includes concepts like social cohesion and social ties, as well as exposure to crime and violence, all of which have been linked to obesity." Also, we used it as a reference when we discussed that, to our knowledge, no studies in Latin America have explored the association between neighborhood social cohesion and obesity (none of the 22 studies included in the provided reference come from Latin America, despite the lack of geographic restrictions in the inclusion criteria). We have also incorporated this reference in several paragraphs in the Discussion – please see enclosed revised manuscript with track changes.

Reviewer: 2
Reviewer Name: Lisa Kakinami and Anna Smyrnova (co-review)
Assistant Professor, Concordia University, Canada (LK)
Graduate student, Concordia University, Canada (AS)

1. Thank you for the opportunity to review Chaparro et al.,'s manuscript entitled, "The association between the neighborhood social environment and obesity in Brazil varies by gender and neighborhood socioeconomic status: a multilevel study". The authors present a well-written investigation into how neighbourhood environments may affect obesity risk. Nevertheless, we have a number of comments and suggestions to the authors for their consideration. In general, key details on many methodological aspects are currently missing from the manuscript, interpretations of classifications (which are based on tertiles) are necessary to give readers a

[1] Pereira de Castro PC, Araujo Nobre A, Ribeiro de Castro IR, Chor D, Harter Griep R, de Oliveira Cardoso L. Does context influece the Body Mass Index of Brazilian workers? Results from the ELSA-Brasil study baseline. *Submitted for publication.*

[2] Boing AF, Subramanian SV. The influence of area-level education on body mass index, waist circumference and obesity according to gender. *International Journal of Public Health* 2015;60:727-736.

[3] Brazilian Ministry of Health, Department of Health Surveillance. [VIGITEL Brasil 2017: Surveillance of risk and protective factors of chronic diseases by telephone survey]. Brasilia, DF 2018. Available at: http://bvsms.saude.gov.br/bvs/publicacoes/vigitel_brasil_2017_vigilancia_fatores_riscos.pdf

better sense of the magnitude of risks and associations, and a more thoughtful discussion/conclusion on what tangibly can be gleaned from this study (rather than a recommendation to ‘reduce the perception of violence’ which is a bit simplistic and insincere) is needed. More specific comments are outlined by section below.

Thank you for your comments, please see detailed responses to all your points below.

Abstract:
2. Some methodological details are lacking. In addition, we feel the conclusion goes beyond what can be concluded from this specific study. In particular, we would caution against using the term “prevent”, as this study reports the association in cross-sectional analysis, and not longitudinal or interventional research.

The abstract is formatted based on the requirements of BMJ Open. Notably, there is not a “Methods” section for us to explain our methods in detail. In the original abstract, we included some methods information in the results section (i.e. “In multilevel logistic regression models adjusted for age, education, and skin color, women…”). At the request of the reviewer, we have now incorporated additional methodological details also in the Results section, when discussing the results related to stratification by neighborhood SES: “When stratified by neighborhood SES – defined based on number of people per household, proportion of children 0-4 years, mean income, and percent of white residents at the neighborhood level – results for social cohesion…”.

As for the conclusion in the abstract, we have added a statement reiterating our cross-sectional findings: “In this civil-servant sample in 6 large cities in Brazil, the neighborhood social environment was associated with obesity among women, but not men.” However, we stand by the second part of the conclusion saying that, based on our findings, neighborhood-level interventions addressing social neighborhood issues <u>may</u> help in the prevention of obesity among women in Brazil. We are not claiming to have conducted such interventions, nor that our results are longitudinal or causal in any way.

Introduction:
3. The connection between obesity and PA, as well as obesity and socioeconomic status is not well established in the introduction. Further details of these are needed to properly ‘set-up’ the research study. (For instance, parts of the information on page 17, line 23 could instead be moved to the introduction).

While we believe that the connection between obesity and PA and obesity and socioeconomic status is well established and hence not needed in the Introduction, we have included additional information explaining the possible mechanisms by which the social neighborhood environment may influence obesity (e.g. PA), as requested by the reviewer.

4. Page 5, Line 22: The introduction alludes to the potential importance of interventions in vulnerable populations. However, as this study is based on civil-servants (which excludes the extremely poor and unemployed) it is not clear how the sample can be considered as a vulnerable population. Some explanation, or rephrasing of that sentence could help.
Thank you for raising this point. Indeed, one of the limitations of this study, as pointed in the paper, is that our sample does not include the unemployed or people living in extreme poverty. Therefore, we have deleted the “particularly among vulnerable populations” from the sentence in question.

5. Page 5, line 29: Some statistics on the recent increase in obesity observed in Brazil (in %) would be helpful here. It would better illustrate why the research question is focused on obesity weight status, rather than on overweight/obesity.

The prevalence of obesity in Brazil in 2016, compared to 2006, has been included at this point in the text, as suggested by the reviewer.

Methods:

6. It would be helpful to provide some statistics on how the current sample differs from the general population (distribution of income, skin color, education attainment etc) to help the readers better understand the generalizablity of study results. Similarly, statistics of the six different cities in this study would be helpful. These tables could be presented in the Appendix.

We believe including statistics on the six cities included in the study is beyond the scope of this manuscript. However, we have incorporated additional information on how the ELSA-Brasil sample varies from the larger Brazilian population in the limitations section as follows: "Moreover, whereas the ELSA-Brasil sample has, on average, a higher income and social class than the residents of the six included cities,[25,44] the ELSA-Brasil sample has a similar prevalence of obesity and obesity-related behaviors (i.e. diet and physical activity patterns) than the Brazilian population at large.[45]"

7. The authors develop their SES measure based on methodology described by Santos et al. (2010). However, Santos et al., defined SES based on four census tract indicators: income, education, persons per household and % in the 0-4 year age bracket. In contrast, this study defined SES not with education, but with percent of white residents. This change in methodology needs to be rationalized and explained. Related to this point, it is always preferable to provide the readers with enough relevant methodological details to understand the study, rather than referring them to an outside publication (such as the Santos et al., study).

The study by Santos et al. (2010) cited in our paper was based on data from the Brazilian Census 2000. The Brazilian Census 2010 did not include education, unfortunately. So, the neighborhood SES variable used in this study (and all other ELSA-Brasil studies using neighborhood-level variables since Census 2010 is available) was created following the same methodology as described by Santos et al. (2010) but with %white residents instead of education.

Our paper does report details regarding the neighborhood SES methodology, including: "neighborhoods were constructed by combining contiguous census tracts with similar sociodemographic composition based on four variables from the Brazilian Census 2010:[21] number of people per household, proportion of children 0-4 years, mean income, and percent of white residents, following the same methodology described by Santos et al. 2010.[20] Neighborhoods were defined with a minimum population size of 5,000 inhabitants, a number deemed appropriate to be able to distinguish between different socioeconomic patterns.[20]" The methodology used by Santos et al. (2010) is rather complex and we believe more details than those provided would be distracting based on the objectives of the current paper. For readers interested in further details, though, the appropriate citation is included.

8. As neighbourhood SES indicators were used for the analysis, more information on the distribution of participants within neighbourhoods is warranted. For instance, it is not clear whether authors used circular or network buffers centered on the participants' locations (to ensure that participants were not between two different neighbourhoods). If not, more details on whether participants were on the 'edges' of two adjacent (but very different SES) neighbourhoods is needed.

We did not use buffers to create the neighborhoods, nor for the neighborhood SES indicator. To create the neighborhoods, each ELSA participant's residential address was geocoded using a double-checked technique. The first one was based on Google Earth mapping databases, allowing us to find a point using geographical coordinates, which was then overlayed into cartographic bases of census tracts polygons. The second one used the Brazilian National Register of Addresses for Statistical Purposes,[4] a database from the Brazilian Institute of Geography and Statistics (IBGE) that allowed us to geocode addresses with each side from street axes precision level to census tracts polygons from the 2010 Brazilian Census. Local neighborhoods were then created using a spatial aggregation method based on SKATER (Spatial 'K'luster Analysis by Tree Edge Removal at TerraView software). This method was used to create

---

[4] Brazilian Institute of Geography and Statistics (IBGE). [National Registry of Addresses for Statistical Purposes]. Available at: https://ww2.ibge.gov.br/home/estatistica/populacao/censo2010/cnefe/default_cnefe.shtm (in Portuguese).

clusters of contiguous census tracts that had a minimum population size of 5000 inhabitants and were homogenous with regard to four socioeconomic indicators from the 2010 Census: income, persons per household, percentage of White population, and proportion of 0-4-year olds.[5] *As the local neighborhood delimitations used at ELSA-Brasil are a set of spatial clusters of contiguous census tracts, with similar SES conditions by definition, all participants that live in the same neighborhood are exposed to similar SES conditions as well as to similar environmental contextual characteristics.*

9. The test-retest values themselves should be provided, rather than a general statement that there was 'good internal consistency'.

This information has been included in the manuscript.

10. The paragraph explaining the quality of scales (Page 8, bottom paragraph) could be moved to right after the sentence when they are first introduced (Page 7, line 45).

This change has been made.

11. As the authors found significant sex differences, were the individual-level scores on social cohesion, perceived safety, and perceived violence also created separately for males and females? If not, rationale in why this was aggregated without adjustment for gender is warranted.

The individual-level scores were calculated separately for each individual participant based on their individual responses to the social cohesion, perceived safety, and perceived violence scales. These individual-level scores were then aggregated to the neighborhood level where participants live so that all participants living in the same neighborhood – men and women – would have the same level of exposure. In other words, anybody in a given neighborhood (of any gender, but also of any age, skin color, individual-level socioeconomic status, etc.) would have the same score for each of the aggregated scales. As displayed in Table R3 below (in response to comment #20), though, neighborhood perceptions don't vary greatly by gender.

Still, as requested by reviewer 3, point #11, we have re-ran the analysis accounting for both individual scores and aggregated neighborhood scores. These results should account for differences in neighborhood perceptions by gender. This information has been incorporated into the paper and is displayed in Tables R4 and R5 below.

12. A brief description and interpretation of the questionnaires' tertiles would be very helpful. For example, what an average value of 10.0 on a social cohesion scale for those in the highest tertile means in terms of social cohesion. (and similarly for perceived safety, and perceived violence)

The social cohesion scale ranges from 5-25 and a higher score indicates a higher social cohesion. A value of 10 in this social cohesion scale does not have a particular meaning, except that it is approximately in the middle of the scale. We could assign it a meaning if it were to be compared to the average value of the sample, which is 17.3 (SD 3.6). In this case, we can say that a score of 10 is approximately half of the mean score for our sample. Because of this and because our three scales have different ranges of responses (which makes interpretation of a random score of e.g. 10 across scales even more difficult), we decided to group scores into tertiles which, by definition, split the population into the 33% bottom scores, the 33% middle scores, and the 33% higher scores. This helps us classify people into relative terms compared to the rest of the sample. Based on this comment, the range, mean values, and SDs for each of the social neighborhood environment variables in each tertile group has been included in Table 1 of the revised manuscript.

13. Similar to the above point, it would be helpful to provide more description of the three SES

[5] Santos SM, Chor D, Loureiro Werneck G. Demarcation of local neighborhoods to study relations between contextual factors and health. *Int J Health Geo.* 2010;9:34.

categories, and how they differ from each other. (A brief table in the Appendix will help the interpretation of the results). For example, what does, “women residing in high and low SES neighbourhoods” mean in terms of the characteristics that these neighbourhoods have? What is the range? In particular, as the study has an inherent limitation of likely excluding the extremely poor population, a range of SES categories provides this important information.

Please see Table R1 below with neighborhood characteristics by neighborhood SES groups. This table is included in the revised version of the manuscript as a supplementary table.

Table R1: Neighborhood characteristics by SES cluster (N=1902 neighborhoods where the sample lived)

	Low SES N=753			Intermediate SES N=550			High SES N=598		
	Mean	**SD**	**IQR**	**Mean**	**SD**	**IQR**	**Mean**	**SD**	**IQR**
Proportion of children 0-4 years	6.85	1.17	6.03-7.62	5.29	0.98	4.62-5.90	3.93	1.04	3.26-4.48
Number of people per household	3.27	0.19	3.15-3.37	3.02	0.18	2.90-3.14	2.59	0.31	2.35-2.81
% White	32.11	15.07	16.31-43.10	59.70	13.79	50.68-69.04	81.24	10.14	74.79-89.11
Median income ($R)	946.56	292.02	733.13-1102.92	1916.65	750.63	1395.50-2209.20	4758.85	2290.00	3003.02-6001.70
Social capital	16.79	2.96	15.25-18.40	17.17	2.62	16.00-18.75	17.45	1.93	16.40-18.62
Perceived safety	8.84	2.61	7.00-10.50	9.35	2.49	8.00-11.00	9.51	1.96	8.25-10.75
Perceived violence	15.68	2.53	14.20-17.50	16.71	2.06	15.81-18.00	17.04	1.67	16.33-18.00

14. It is not clear why the educational attainment categories of “less than primary” and “primary” were both retained in the regressions. Is this a meaningful comparison group to describe? Were there in fact significant differences noted between these two groups? Otherwise, the ‘non-linear trends’ shown

between these different educational categories in the results (eg: table 2, model 3, greater risk was found for those with a secondary education, than with a primary education) needs to be mentioned in the discussion.

Our main goal was to investigate the association between different social neighborhood environment characteristics and obesity, while adjusting for potential confounders; namely, age, education, and skin color. Changing the classification of education does not change the results (ORs, 95%CIs) obtained for our main predictors (i.e. neighborhood social cohesion, perceived safety, and perceived violence). Discussing the association between education on obesity – a well-established association – was not part of our objectives, despite us showing the numbers in the table for interested readers. Still, based on this comment we have now updated all analyses and tables to merge the less than primary and primary education categories.

15. Page 10, line 17: “neighborhood level scores of our three neighborhood variables were reconverted into tertiles within each neighborhood SES category” – were any significant differences in scores noticed? Could you provide any figures/table about the distribution of SES categories and distribution of scores within each SES category? (it’s interesting to see if, for example, score for violence was ‘XX times less’ in high SES than in low SES)

Please see Table R1 above, which also includes this information.

Results:
16. It’s not clear from Table 1 whether the distribution of age, education and skin color (etc) differ by sex. As the paper makes the argument for important sex differences, further information on these sex differences should be assessed and provided for the readers.

Average age did not vary between men (mean 52.0 years, SD 9.34) and women (mean 51.9n years, SD 8.84). Overall, women in the sample had a higher education than men and were more likely to be black (Table R2). A summary of this information has been added in the text of the revised manuscript, when describing sample characteristics.

Table R2: Differences in sample characteristics by gender

	Women %	**Men** %
Education		
Primary or less	9.50	16.23
Secondary	36.29	33.46
University	54.21	50.31
Skin color		
Black	18.61	14.33

Brown	27.80	30.87
White	53.59	54.80

17. Were there any correlations between the various neighbourhood scores with one another? Is there any concern that they may be multicollinear with one another and so are presenting 'redundant' information in some of the regression models? This may have implications on including them in the same (multilevel) model, such as presented in table 5.

The models presented in Table 5 do not control the estimated associations of each perception variable for the other two perception variables. Each of the models (for social cohesion, perceived safety, and perceived violence as independent outcomes) have been adjusted by age, education, and skin color only, in tune to the previously presented analysis. The results are just presented all in one table due to space limitations. This has been clarified in the table.

There is a low correlation between the scores for the social cohesion scale and both the perceived safety (Pearson correlation coefficient 0.24) and perceived violence (0.26) scales. There is a moderate correlation between the perceived safety and perceived violence scales (0.46) These correlation coefficients are similar for women and men. This information has been included in the Methods section when the scales are first discussed.

We would not have issues of multicollinearity since, as just explained, we have not put these variables in the same model. But these low-to-moderate correlation coefficients do highlight that we are measuring different (albeit related) concepts within the social neighborhood environment.

18. The article cited by authors (Sampson et al. 1997) suggested that "a measure of collective efficacy (defined as social cohesion among neighbors…) … is negatively associated with variations in violence" – was this association checked in this study?

Please see response to comment #17.

Discussion:

19. As mentioned previously, some explanation of the different SES tertiles will help with interpreting what 'high' and 'low' SES neighbourhoods mean (eg: Page 17, line 20)

Neighborhood SES classifications are not based on tertiles but on principal components and cluster analyses based on 4 neighborhood-level characteristics: number of people per household, proportion of children 0-4 years, median income, and % white residents. Please see Table R1 for characteristics of neighborhoods based on SES categories, which is now included in the paper as a Supplementary Table.

20. The authors mention that gender related differences in neighbourhood perceptions have been previously reported in the literature. Did the authors also find this to be the case in this study?

Individual-responses to the social cohesion scale, the perceived safety, and the perceived violence scales did not differ by gender in any meaningful way (Table R3). So, both men and women rate their neighborhoods in a similar manner, but for women, these ratings affect their obesity risk, whereas this is not the case for men. This information has been added to the Discussion as follows: "We found that the neighborhood social environment only affects obesity risk among Brazilian women and not men, even *though there were no gender differences in the social cohesion, perceived safety, and perceived violence average individual scores."*

Table R3: Average individual scores for neighborhood social cohesion, perceived safety, and perceived violence by gender

	Women Mean (SD)	**Men** Mean (SD)
Social cohesion	17.52 (3.67)	17.14 (3.56)
Perceived safety	9.38 (3.20)	9.67 (3.14)
Perceived violence	16.78 (2.82)	16.82 (2.86)

21. Even though perception of violence and 'actual' level of violence (based on crime statistics, for example) are likely associated, we'd caution the authors against using these terms in an exchangeable way, especially since objective crime measures were not incorporated. Some neighbourhood interventions could be proposed to reduce the actual violence, but violence perception depends also on many other personal factors. As the authors cannot obtain objective measures of crime, references linking the association between actual violence and perceived violence would be warranted here (Page 20, first paragraph in particular).

We have been very careful about always referring to our findings as "perceived violence" and never claimed to be discussing objective violence; we do not believe to have used the terms interchangeably in the paper. We did include in the discussion potential interventions to reduce levels of violence (not perceived violence). We have now expanded that section to avoid confusion, as follows: "Our results suggest that neighborhood interventions to increase social cohesion and decrease violence perceptions may prevent obesity among women in Brazil. Effective neighborhood interventions designed to reduce violence may include the cleaning and greening of vacant lots, as well as the reduction of alcohol availability. Though the effect of these kinds of interventions on <u>perceived</u> violence is unknown, research suggests that <u>fear of crime</u> may be negatively influenced by neglected and run-down neighborhood spaces."

22. The conclusions could use a bit more rumination. We feel it is an overgeneralization to conclude that increasing social cohesion and violence perception could be protective against obesity. This type of conclusion is a bit simplistic, and may be misconstrued as flippant and insincere. A more thoughtful conclusion would incorporate more tangible suggestions on what policy, researchers, or communities can do to try to help combat this problem together.

We politely disagree with this comment. Even though our results are novel for the Latin American context, there are in line with research conducted in the U.S. and other high-income settings. Namely, social neighborhood characteristics have an impact on obesity risk among its residents, more so among women than men. We are suggesting possible interventions at the neighborhood level that could address the social characteristics we found to be associated with obesity among Brazilian women (social cohesion and perceived violence). These are tangible suggestions that could be implemented at the community-level. We further suggest in the conclusions that intervention researchers test the proposed interventions addressing social cohesion and perceived violence, and to further explore the gender differences observed in this study and others in different contexts. These are also tangible suggestions for researchers.

<u>Tables:</u>
23. Table 1: The percentages should be the 'column' percentages, and not the 'row percentages'. IE, it should be the percent of women who were 'not obese' compared to the men who were 'not obese', rather than the percent of all women who were 'obese' compared with the percent of all

women who were ‘not obese’. Statistical comparisons should be done, and p-values should be in the table.

Table 1 shows the characteristics of the sample (total, Column %) as well as the characteristics of those who are obese (and not). We politely disagree with the reviewer’s comment indicating that we should present column percentages. What we are trying to display is, e.g., how the prevalence of obesity varies for women (23.4%) vs. men (19.9%) and the same for all the other included characteristics. If the reader would like to get other comparisons as those suggested up by the reviewer, all the necessary numbers for calculation are included in the table. In addition, since we are following STROBE guidelines, statistical comparisons of sample characteristics are not included:

STROBE guidelines (page 1643)[6]: “Inferential measures such as standard errors and confidence intervals should not be used to describe the variability of characteristics, and significance tests should be avoided in descriptive tables.”

Minor corrections:

24. Abstract, Line 43: High SES and low SES neighbourhoods are mentioned, yet the use of tertiles is not described. The total range is not clear. Possible suggestion is to use “highest SES out of three tertiles” or “high and low tertiles”.

Neighborhood SES classifications are not based on tertiles but on principal components and cluster analyses based on 4 neighborhood-level characteristics: number of people per household, proportion of children 0-4 years, median income, and % white residents. This information has been added to the abstract as explained in the response to comment #2 above.

25. Page 7, line 38: typo – (BMI **≥** 30)
Table 2: Typo in the title (BMI ≥ 30)
Table 3: Typo in the title (BMI ≥ 30)
Table 4: Typo in the title (BMI ≥ 30)
Table 5: Typo in the title (BMI ≥ 30

Thank you for catching these. All typos were fixed.

Reviewer: 3
Reviewer Name: Ruben Brondeel
Ghent University - Department of Movement and Sports Sciences

1. This article investigates important relations between neighborhood characteristics and BMI, in the context of social inequalities (gender and SES). The results are based on a very interesting data set, in terms of its origin, of its size and – despite the inclusion of only civil servants – of its composition. However, I have some comments on the methods. The analysis can be improved, which would make the results more trustworthy.

Thanks for this comment, we are pleased you find our study interesting. Further comments below.

Major comments:

2. The authors took decisions in the model building that are not motivated in the article and, in my opinion, not appropriate.

> The dependent variable, BMI, was dichotomized. This decision is not motivated, and leads to a loss of information and power.
> Dawson, Neal V., and Robert Weiss. "Dichotomizing continuous variables in statistical analysis: a practice to avoid." (2012).

We agree that statistical power is lost when continuous variables are dichotomized. However, in the case of obesity, the dichotomization of BMI – a notoriously non-linear variable – has clinical and public health

[6] Vandenbroucke JP et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Med* 2007;4(10; e297):1629-1654.

importance. If we were to use BMI as the outcome instead of obesity, our results would be harder to interpret for a public health audience (our target). For example, we could find that those in the lowest tertile of the neighborhood social cohesion, compared to those in the highest, would have an X kg/m^2 higher BMI. What does this result really tell us? Unless this person is crossing BMI categories (moving between the underweight, normal weight, overweight, or obesity) a X-unit increase may be meaningless (e.g. moving between a BMI of 21 to 22 may not matter but moving from 24 to 25 does have a meaning). Therefore, we stand behind our dichotomization of BMI into obesity.

3. The independent variables were categorized in tertiles. Similar to dichotomizing the dependent variable, this has considerable consequences.

Mabikwa, Onkabetse V., et al. "Assessing the reporting of categorised quantitative variables in observational epidemiological studies." BMC health services research 17.1 (2017): 201.
Collins, Gary S., et al. "Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model." Statistics in medicine 35.23 (2016): 4124-4135.
Kahan, Brennan C., et al. "A comparison of methods to adjust for continuous covariates in the analysis of randomised trials." BMC medical research methodology 16.1 (2016): 42.

Please see response to reviewer #2, comment #12 for our justification of using tertiles.

4. The model presented in Table 5 includes the 3 neighborhood perception variables, controlling the estimated associations of each perception variable for the other two perception variables. In Table 2 to Table 4, the associations of the 3 variables are tested separately. The two different approaches are not motivated. I suggest to use the modeling strategy from the model in Table 5 throughout the article.

The models presented in Table 5 do not control the estimated associations of each perception variable for the other two perception variables. Each of the models (for social cohesion, perceived safety, and perceived violence as independent outcomes) have been adjusted by age, education, and skin color only, in tune to the previously presented analysis. The results are just presented all in one table due to space limitations. This has been clarified in the table.

5. Moderator effects are tested in sub-group analyses. Introducing interaction terms in the regression model would result in less estimates and therefore in more precision in the estimations.

While the reviewer is correct in pointing out that the models with interactions provide less estimates and more precision, there are also more difficult to interpret than those obtained from stratified analysis. Therefore, we decided to keep our analysis as presented, stratified by neighborhood SES.

6. There is no motivations for the model building in Tables 2 to 4 (i.e. adding variables in the different models). This type of model building could be used as a rudimentary way of understanding mediation mechanisms. However, there are no such objectives in the article. And if there were, the sequence of introducing the variables would be different, with the perceptions being introduced in the final model.

Since we are requested to follow STROBE guidelines (please see "Main results: (a)" in STROBE statement), we are displaying unadjusted (or, in our case, minimally adjusted by age; Model 1) models and models adjusted by our included confounders (Model 3). We agree that perhaps Model 2 is unnecessary, so we have removed it from Tables 2-4. Now former Model 3 is Model 2 in the revised manuscript.

7. The variables 'perceived safety' and 'perceived violence' are highly related, not only by concept but also by construct. One of the three indicators of 'perceived safety' is 'Violence is a problem in my neighborhood', which could be clearly an indicator 'perceived violence'. The authors should give more information (e.g. correlation coefficients between 3 perception variables) and explanation (e.g. origin of scales) to motivate the distinction between the two constructs. Alternatively, a variable can be removed from the analysis.

Please see response to reviewer #2, comment #17. We have a moderate correlation between perceived safety and perceived violence scores (Pearson correlation coefficient=0.46). Our interpretation of this correlation coefficient is that, although related, we are measuring different concepts within the social neighborhood environment. This information has been included in the paper in the Methods section when discussing the three scales.

Minor comments:
Introduction:
8. Social environment is loosely defined. What other aspects are often included in social environment, next to those included in this article? And how was the choice for 'social cohesion', 'perceived safety' and 'perceived violence' made?
To our knowledge, there is no established definition of what the social environment entails. As we say in the introduction, the social environment refers to the social interactions that occur in the neighborhood between neighbors. This could include many more things than those included in this study, like community organizing, participation in social clubs, community kitchens, etc. The neighborhood social aspects mentioned in the introduction and those included in this study (social cohesion, safety, violence) have been those most often studied in reference to obesity. This is primarily because there are plausible explanations linking these neighborhood social characteristics and obesity (as originally included in the discussion and now also in the introduction because of this comment and one from reviewer #2) and because there are existing scales and other sources of data (e.g. crime reports) to quantify these social characteristics.

9. Why is the adjective 'perceived' not used for 'social cohesion'? I do not see a difference with 'safety' or 'violence'.

Social cohesion is an inherently subjective concept. Safety and violence could be measured in an objective manner, though, through traffic reports, crime reports, etc. Since we do not have access to objective-related safety/crime data, we have made the distinction that the safety/crime variables included in the study are based on individuals' perceptions.

Methods: (note that some of the comments below might become irrelevant if the major comments are addressed)
10. (p7 - line 26) The mean can be easily calculated from the data in the line above. It would be more informative to report percentiles (e.g. the median, quartiles, minimum and maximum, …), especially because a skewed distribution can be expected.

As requested by the reviewer, this information has been included.

11. (p9 – line 3) The aggregate of the perceived measures were used as neighborhood characteristics. I suggest to investigate both the individual and aggregated variables in the same regression model, to test the neighborhood effects above and beyond the individual effects.

Thank you for this suggestion. We have re-ran all of our analyses adjusting by individual-level scores on social cohesion, perceived safety, and perceived violence scales. Interestingly, all of our findings are robust to adjusting for individual-level scores (where the respective aggregated values are the outcome), as displayed in Tables R4 and R5. Our Methods, Results and Discussion have been updated to reflect this change in the analysis.

12. Since there is a strong theoretical difference in perceived safety/violence between women and men, should you use gender-specific aggregated measures?

The individual-level scores were calculated separately for each individual participant based on their individual responses to the social cohesion, perceived safety, and perceived violence scales. These individual-level scores were then aggregated to the neighborhood level where participants live so that all participants living in the same neighborhood – men and women – would have the same level of exposure. In other words, anybody in a given neighborhood (of any gender, but also of any age, skin color,

individual-level socioeconomic status, etc.) would have the same score for each of the aggregated scales. As displayed in Table R3 above (in response to reviewer #2, comment #20), neighborhood perceptions don't vary greatly by gender.
Still, as requested by the reviewer and as included in our response to point #11 above, we have re-ran our analysis further adjusting by individual-level scores. This further adjustment should take care of the minimal gender differences that exist in neighborhood perceptions.

13. What are the Intra-class-correlation coefficients for the three perception variables? You could estimate these in multilevel models with each of the variables (before aggregation) as a dependent variable, a random effect for neighborhood, and no independent variables. Even though ICC is not a perfect measure, it would be an indication of how appropriate the aggregated measures are.

We believe that by adjusting by both individual and aggregated scores, as suggested by the reviewer, this check becomes unnecessary.

14. (p9 – line 29). The construction of neighborhood SES is not motivated. Why was there no cluster analysis directly on the 4 indicators? Why was there a cluster analysis at all? With the high number of participants included in the study, there is no need for variable reduction, and the 4 SES variables could have been used separately. Using the 4 variables in their original form (in combination with interaction terms instead of the sub-group analyses), would be more informative and would make it easier to reproduce the analyses. The sequence of variable reduction methods make the resulting SES variable very specific to the dataset at hand.

Our interest was not to estimate the effect of each of these SES-related variables on the association between the social environment and obesity. Rather, we wanted to see if the observed associations between the social environment and obesity varied by neighborhood SES. As such, we were interested in a summary measure of SES that was specific to the Brazilian context, which is what we created.

As pointed out by the reviewer, we could have done the cluster analysis directly on the four variables instead of running principal components and then a cluster analysis. However, the resulting clusters would have been more difficult to interpret as we would have had to interpret the variability of 4 variables instead of the two principal components.

Table R4: Results from the gender-stratified multilevel logistic regression models predicting obesity (BMI≥30 kg/m2) by neighborhood social cohesion, by neighborhood perceived safety, and by neighborhood perceived violence independently. Adjusted by age, education, and skin color; and individual social cohesion score (for neighborhood social cohesion model), or individual perceived safety score (for neighborhood perceived safety model), or individual perceived violence score (for neighborhood perceived violence model)

	Social cohesion OR (95% CI)		**Perceived safety OR (95%CI)**		**Perceived violence OR (95%CI)**	
	Women	**Men**	**Women**	**Men**	**Women**	**Men**
Individual scores						
Social cohesion	1.00 (0.98-1.02)	0.99 (0.97-1.01)				
			1.00 (0.94-1.02)	1.00 (0.98-1.03)		

Perceived safety					0.98 (0.96-1.00)	0.99 (0.96-1.02)
Perceived violence						
Neighborhood social characteristic						
Lowest tertile	**1.26 (1.00-1.58)**	0.90 (0.70-1.16)	1.14 (0.92-1.41)	1.04 (0.82-1.32)	**1.27 (1.01-1.59)**	1.06 (0.82-1.36)
Middle tertile	1.08 (0.91-1.29)	0.94 (0.78-1.15)	0.95 (0.80-1.14)	0.98 (0.80-1.19)	1.03 (0.86-1.23)	0.99 (0.82-1.20)
Highest tertile	1.00	1.00	1.00	1.00	1.00	1.00
Age	**1.02 (1.01-1.03)**	1.00 (1.00-1.01)	**1.02 (1.01-1.03)**	1.00 (1.00-1.01)	**1.02 (1.01-1.03)**	1.00 (1.00-1.01)
Education						
Up to primary	**1.47 (1.14-1.89)**	1.10 (0.84-1.44)	**1.50 (1.16-1.93)**	1.10 (0.84-1.43)	**1.44 (1.11-1.86)**	1.09 (0.83-1.42)
Secondary	**1.48 (1.26-1.73)**	1.10 (0.90-1.34)	**1.49 (1.27-1.74)**	1.10 (0.90-1.33)	**1.44 (1.23-1.68)**	1.09 (0.89-1.32)
University	1.00	1.00	1.00	1.00	1.00	1.00
Skin color						
Black	**1.82 (1.50-2.21)**	**1.40 (1.08-1.82)**	**1.83 (1.51-2.23)**	**1.40 (1.08-1.81)**	**1.79 (1.47-2.18)**	**1.39 (1.07-1.81)**
Brown	**1.35 (1.13-1.61)**	1.11 (0.91-1.36)	**1.36 (1.14-1.63)**	1.11 (0.91-1.36)	**1.34 (1.12-1.61)**	1.11 (0.90-1.36)
White	1.00	1.00	1.00	1.00	1.00	1.00

Table R5: Results from multilevel logistic regression models predicting obesity (BMI≥30 kg/m2) by neighborhood social cohesion, by neighborhood perceived safety, and by neighborhood perceived violence independently, stratified by neighborhood SES. Adjusted by age, education, and skin color; and individual social cohesion score (for neighborhood social cohesion model), or individual perceived safety score (for neighborhood perceived safety model), or individual perceived violence score (for neighborhood perceived violence model)

	High SES	**Intermediate SES**	**Low SES**
	OR (95% CI)	**OR (95%CI)**	**OR (95%CI)**

	Women	Men	Women	Men	Women	Men
Social cohesion						
N	2799	2144	1882	1371	1410	1268
Lowest tertile	**1.52 (1.10-2.10)**	1.01 (0.71-1.43)	0.90 (0.59-1.37)	0.96 (0.62-1.50)	1.42 (0.92-2.18)	0.91 (0.55-1.50)
Middle tertile	1.07 (0.81-1.42)	0.99 (0.73-1.34)	1.07 (0.78-1.48)	0.82 (0.57-1.18)	1.00 (0.72-1.41)	0.85 (0.57-1.28)
Highest tertile	1.00	1.00	1.00	1.00	1.00	1.00
Perceived safety						
N	2804	2148	1882	1371	1413	1269
Lowest tertile	1.03 (0.73-1.44)	0.96 (0.67-1.37)	0.82 (0.55-1.22)	1.06 (0.68-1.66)	1.30 (0.84-2.00)	1.12 (0.68-1.84)
Middle tertile	0.97 (0.72-1.30)	0.85 (0.62-1.15)	0.85 (0.61-1.18)	0.96 (0.66-1.40)	1.24 (0.89-1.73)	0.94 (0.63-1.40)
Highest tertile	1.00	1.00	1.00	1.00	1.00	1.00
Perceived violence						
N	2799	2139	1875	1369	1412	1269
Lowest tertile	0.98 (0.70-1.37)	1.16 (0.79-1.71)	1.25 (0.83-1.89)	1.01 (0.64-1.60)	**1.84 (1.17-2.91)**	0.98 (0.57-1.68)
Middle tertile	0.87 (0.65-1.15)	1.06 (0.78-1.42)	1.02 (0.72-1.44)	0.83 (0.56-1.22)	**1.66 (1.16-2.37)**	1.00 (0.67-1.50)
Highest tertile	1.00	1.00	1.00	1.00	1.00	1.00

15. (p10 – line 5). Are the ELSA sites the 6 cities? If so, you could consider introducing the variable as a fixed effect instead of a random effect. This would provide more information and simplify the analyses.

Yes, the ELSA sites are the 6 cities in which participants lived. These 6 cities are very different from each other in sociodemographic composition, also in geographic location and everything that that entails. As such, we believe including them as a random effect is more appropriate. Indeed, if we test whether the variances of the neighborhood-level random intercept and the ELSA sites random intercept are zero (i.e. testing the significance for these random effects), we find that not to be the case. The table below shows an example of this finding for the multilevel logistic regression model for women with social cohesion as the predictor:

Tests of Covariance Parameters Based on the Residual Pseudo-Likelihood					
Label	**DF**	**-2 Res Log P-Like**	**ChiSq**	**Pr > ChiSq**	**Note**
var(centroa)=0	1	28045	12.63	0.0002	MI
var(a_geocodviz(centroa))=0	1	28036	3.70	0.0272	MI

The variable "centroa" is the ELSA site and the variable "a_geocodviz" is the neighborhood. The p-values<0.05 indicate that we reject the null hypothesis that the random intercepts are zero and, hence, we need to include these random effects in our analyses.[7]

Results:

16. p10 – line 30: The second 'sample' should be deleted.

This change has been made.

17. p10 – line 33: Use explicitly the age boundaries for 'middle aged people', because the term 'middle-aged' has many different uses.

This has been included.

18. P11 – line 39: The result presented here indicates the lack of a moderation effect (same effect, independent of the SES status). An overall dose-response association (i.e. a linear association) could be derived from Table 4, or a group-specific dose-response association could be derived in Table 5 for each group specific.

We apologize for the confusion. We meant that our findings suggested a group-specific dose-response association; in this case, for women in the low SES category, for the association between perceived violence and obesity. This has been clarified in the text.

19. Table 5: What are the sample sizes in the 6 different samples? For women in low SES neighborhoods, there are some effect sizes which are not statistically significant, but they are as high as some estimates that were statistically significant in the other tables. The reader cannot derive if this is due to a lack of power related to the sample size, or to a higher standard deviation.

Sample sizes have been included in the revised version of Table 5.

Discussion and conclusion:

20. In the strobe statement, you mention that the study is explorative. However, the discussion section (nor the methods) does not reflect this. For example, p18 – line 11, the results from two out of three variables (and only one in the subgroup analyses in Table 5) are generalized to 'the neighborhood social environment'. I suggest that the discussion stays a little closer to the results, and indicates what is missing in the results (e.g. other dimensions of social environment, objective measures of social environment).

We have included a hypothesis at the end of the Introduction section. In addition, we have revised some of the statements in the Discussion to be outcome-specific and not generalized to the neighborhood social environment.

STROBE statement:

[7] Zhu M, SAS Institute Inc. Analyzing multilevel models with the GLIMMIX Procedure. Paper SAS026-2014. Available at: https://support.sas.com/resources/papers/proceedings14/SAS026-2014.pdf

21. Title: the term ‘multilevel’ rather reflects the type of analysis than the study design. ‘Cross-sectional study’ or ‘cohort study’ would be more in its place.

The title has been changed to “The association between the neighborhood social environment and obesity in Brazil varies by gender and neighborhood socioeconomic status: a cross-sectional analysis of the ELSA-Brasil study.”

22. Descriptive data: the comment that complete data is used, is not an alternative to describing the (relatively high number of) participants with missing data. The reader needs to understand why some people had missing data and were therefore left out the analysis.

A breakdown of missing data has been included in the revised version of the STROBE statement as follows:
WOMEN max sample = 8218 (all have age and education)
8218 – 395 with missing skin color = 7823
7823 – 3 missing obesity = 7820
7820 – 1713 missing neighborhood (and hence, all neighborhood values) = 6107 → ANALYTICAL SAMPLE

MEN max sample = 6887 (all have age and education)
6887 – 320 with missing skin color = 6567
6567 – 3 missing obesity = 6564
6564 – 1773 missing neighborhood (and hence, all neighborhood values) = 4791 → ANALYTICAL SAMPLE

VERSION 2 – REVIEW

REVIEWER	Lisa Kakinami and Anna Smyrnova (co-review) Assistant Professor, Concordia University, Canada (LK) Graduate student, Concordia University, Canada (AS)
REVIEW RETURNED	03-Apr-2019

GENERAL COMMENTS	Although the authors have addressed many of the smaller issues/concerns that were previously raised, there are a few major items that we feel are inadequately addressed in the manuscript, as well as in the response to reviewers. These comments are below. Major comments: 1. If the paper is intended to investigate whether a person's perception of their neighbourhood may be associated with their obesity risk, it is unclear why the authors aggregated the scores so that every person living in the same neighbourhood would have the same 'exposure'. We feel that the authors did not adequately address this concern when it was previously raised. Although the authors aimed to account for individual and neighbourhood scores separately, it is unclear why the authors could not use the neighbourhood scores based on the census variables, and the individual scores based on the perceptions. To use the aggregated neighbourhood scores (which were not sex-stratified) would bias their results and is counter to their objective of testing for sex differences.

	2. The authors provide a lot of important information in their 'response' pertaining to how they identified clusters of contiguous census tracts. Based on their description, it appears that the authors assumed that people living in the same polygon (regardless of if they were in the edge, center, etc) had by definition, the same SES and "environmental contextual characteristics". However, weighting these exposures and/or accounting for differences in these characteristics for the households in the center (vs the edge) is common practice and should be done. 3. More details on identifying the neighbourhood principal components and clusters are needed in the methods. As both of these methodologies are very data-driven, a more thorough description of these methods, and an acknowledgement of how the results may differ widely in a different sample is needed. 4. The use of tertiles may be especially problematic when stratifying by SES cluster, as the difference in the means (based on supplementary table S1) are all quite small. The authors should be mindful of whether any statistical significance is indeed meaningfully different. The authors argue for the use of the tertiles as the three scales had different ranges and were not easily comparable to one another. However, the authors could have used other methods (such as standardization, or using defined cut-offs based on the literature) to create meaningful comparison groups. 5. Building on the previous items, a more thoughtful discussion of the limitations of the research is needed. Any one of the previous comments would be considered a major limitation in a research study, and more transparency in what the study can (and cannot do) is needed. 6. Related to the previous item, there is concern that the authors made conclusions which were not justified by the results. In the discussion, the authors describe how the study results can be extrapolated to future interventions. This feels like a big leap to make based on the study that was conducted, and the authors should not overextend or extrapolate beyond what their study results demonstrate. Minor comments: 7. Introduction: The necessity of a sex-stratified analysis/objective of the paper is not well justified with the current introduction. 8. Methods: As previously mentioned in the last round of comments (item 7), the authors create neighbourhoods based on methodology by Santos et al. However, an important difference between this paper and the methodology by Santos is the use of 'percent of white residents' in identifying neighbourhoods. We feel this is quite a departure in methodology from that by Santos et al., and should be justified. 9. Table 1 should include p-values 10. Regression tables should include the sample sizes

REVIEWER	Ruben Brondeel Ghent University
REVIEW RETURNED	25-Mar-2019

GENERAL COMMENTS	General:

	I appreciate the work that has been done by the authors. However, I feel some of the reviewers' remarks have been not been answered completely. For the first two comments, I acknowledge that these methods (categorizing dependent and independent variables) are common business in the literature, and that the other reviewers did not mention this point. Response to response: 2. BMI categories 'normal weight', 'overweight' and 'obese' have been criticized by many authors for being arbitrary. The argument that differences between 22-23 are different from 24-25 has no real scientific basis. If the link with these categories needs to be made to help interpretation by clinicians, this could be done in a sub-analysis. Also, non-linearity can be handled using continuous variables 3. No justification for the tertiles can be found in comment 12 to reviewer 2. As in the comment about dichotomizing BMI, if specific contrasts need to be made - which doesn't seem to be the case in the article since the results are not interpreted in such a way - this can be done in a sub-analysis. 5. There is no reason that an interaction effect would be more difficult to interpret. Interaction terms estimate the same differences as those in sub-group analyses, so it's just about representing the results in the correct way. Also, interaction terms would provide an insight in the differences in results between, for example, women and men. 7. My remark wasn't about the Pearson correlation. There is, in my opinion, a problem with the (face) validity of the items. 9. (Some aspects of) social cohesion could be measured objectively. For example, with methods counting the number of contacts a person has during the day (e.g. smartphone Bluetooth data). 14. PCA searches by definition for perpendicular dimensions. So clustering them in 3 categories and then naming them 'low', 'middle' and 'high', as if they are on 1 dimension, does not make sense. From this methodology, you would expect categories like 'low-low', 'low-high', 'high-low', 'high-high', in its most basic form. 15. I believe there is a some misunderstanding in the use of a random vs fixed effect. The test presented in the comments is not relevant to this, since it tests a model with random effect to a model without any effect. Also, any variable (e.g. age groups) that explains some variance could give you a significant test result. The fact that the centers are very different, is no motivation for a random effect vs a fixed effect either. Both types of effects would adjust for the relatedness between participants in the same center. 'Centroa' could be introduced as both random or fixed term since it has only 6 categories and each category includes many participants. Fixed effects have the advantage that you get interpretable coefficients, and there are less assumptions in the calculation of the coefficients. A random effect really becomes necessary if there are so many categories (with usually relatively few participants), a fixed effect would no longer be interpretable. New comments: Introduction: - The hypothesis are not completely based on the introduction. There are no arguments to include gender and neighborhood SES

	as moderators to the model, or why other variables are not considered as moderators (e.g. age). Discussion: - In the discussion, it becomes clear that food and physical activity neighborhood variables were available. These variables should have been included in the model, or it should be argued why they are not. At first sight, they seem to possibly confound the reported associations. - P20 line 25: it doesn't really matter that there are no gender differences in the social environment variables, the associations could still result from different mechanisms. (difference between means and covariances). - P20 line 46: as in previous comment, the first reason is not really relevant. The same or different perceptions could both result in similar or different associations. (see p22 line 3, for an example of this). - P22 line 48: as mentioned by another reviewer, the causality is put too strong here. There should be at least a more thorough discussion on the exact mechanism that social cohesion influences obesity. Strengths and limitations: - P23 - line 20: Could you strengthen the argument by providing some comparison between the general population and the sample? Since poverty and poor neighborhoods is very significant for the reported associations, it would be good to have some more information on this (e.g. percentage living in low/middle/high SES neighborhoods). - P23 – line 27: it does not really matter that the prevalence of obesity is about the same in the sample as in the Brazilian population. It is somewhat encouraging, but it is possible that within two groups with the same prevalence of obesity, there are two totally different mechanisms that lead to obesity (see remarks above). Minor remarks: P6 line 17: … to work … to be protective P23 lines 44 – 51: it is great to refer to selective residential mobility. Is it possible to rewrite this argument to make this clearer to the reader. P24 line 18: replace 'living in poverty' by 'living in poor neighborhoods'.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Reviewer Name: Lisa Kakinami and Anna Smyrnova (co-review)
Assistant Professor, Concordia University, Canada (LK)
Graduate student, Concordia University, Canada (AS)
Although the authors have addressed many of the smaller issues/concerns that were previously raised, there are a few major items that we feel are inadequately addressed in the manuscript, as well as in the response to reviewers. These comments are below.

Major comments:

1. If the paper is intended to investigate whether a person's perception of their neighbourhood may be associated with their obesity risk, it is unclear why the authors aggregated the scores so that every person living in the same neighbourhood would have the same 'exposure'. We feel that the authors did not adequately address this concern when it was previously raised. Although the authors aimed to account for individual and neighbourhood scores separately, it is unclear why the authors could not use the neighbourhood scores based on the census variables, and the individual scores based on the perceptions. To use the aggregated neighbourhood scores (which were not sex-stratified) would bias their results and is counter to their objective of testing for sex differences.

The objective of the study was to investigate the contribution of neighborhood-level characteristics to obesity risk, while controlling for individual-level variables. As such, different sources of information, each adequate to the level of exposure involved in the development of obesity, are incorporated into the analytical models. In this way, both individual and contextual factors are considered. The neighborhood variables are based on self-reported measures collected at the individual level, but represent characteristics related to the construct of a contextual attribute of the environment in which the participant is embedded (i.e., their neighborhood).

Self-reported neighborhood perception scales are tools developed for the measurement of contextual dimensions and, even though they are collected as individual responses, they are designed to be analyzed aggregated as contextual attributes of respondents' shared living environment. These scales were developed to capture individuals' perceptions of the social and physical aspects of their neighborhoods, which data from administrative sources (e.g. census) cannot fully capture, thus allowing for the use of detailed information more closely associated to the outcomes of interest.[8,9]

These instruments have been widely used, mainly in the past three decades, in epidemiological studies following multilevel approaches for the identification of contextual factors and their contribution to individual health outcomes. These instruments were developed initially by Chaskin (1997),[10] *Sampson et al. (1997),*[11] *Raudenbusch and Sampson (1999),*[12] *Sampson (2003),*[13] *Echeverria et al. (2004),*[14] *and Mujahid et al. (2007),*[1] *and later translated, adapted, and validated by Santos et al. (2013)*[15] *for application in ELSA-Brasil. Analyses of these data using multilevel models are appropriate to take into account the interdependence in the information collected at the individual level of those who occupy a shared space with similar environmental characteristics, ensuring that the neighborhood scores – such as social cohesion and security – are treated as an attribute of the context. It is thus assumed in the analysis that individuals living in the same area (neighborhood) are exposed to similar environmental constraints (whether of the physical environment, built, or the natural environment).*

2. The authors provide a lot of important information in their 'response' pertaining to how they

---

[8] Mujahid, MS; Diez-Roux, AD; Morenoff, J;Raghunathan, T. (2007) Assessing the measurement properties of neighborhood scales: From psychometrics to ecometrics. *Am J Epidemiol.* 165: 858-867.
[9] Diez-Roux, A & Mair, C. (2010) Neighborhoods and health. *Ann NY Acad Sci.* 1186:125-145.

[10] Chaskin, R. J. (1997) Perspectives on neighborhood and community: a review of the literature. *Soc Ser Rev.* 71(4): 521-527.
[11] Sampson, R.J., Raudembush, S.W. & Earls, F. (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science* 277(5328): 918-24.
[12] Raudenbush, S. W. & Sampson, R. (1999) Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Soc Method* 29(1): 1-41.
[13] Sampson, R. J. (2003) The neighborhood context of well-being. *Persp Biol Med* 46(3-suppl): S53-S64.
[14] Echeverria, S. E.; Diez-Roux, A. V. & Link, B. G. (2004) Reliability of self-reported neighborhood characteristics. *J Urban Health*. 81(4): 682-701.
[15] Santos, SM, Griep, RH, Cardoso, LO, et al. (2013) Cross-cultural adaptation and reliability of measurements on self-reported neighborhood characteristics in ELSA-Brasil. *Rev Saúde Pública* (47), Supl 2:122-130.

identified clusters of contiguous census tracts. Based on their description, it appears that the authors assumed that people living in the same polygon (regardless of if they were in the edge, center, etc) had by definition, the same SES and "environmental contextual characteristics". However, weighting these exposures and/or accounting for differences in these characteristics for the households in the center (vs the edge) is common practice and should be done.

Yes, we assumed that people living in the same polygon (regardless of if they were in the edge, center, etc.) had the same SES and environmental contextual characteristics. This methodology is the same to that used by other large studies in the field that have been validated and replicated including the Project on Human Development in Chicago Neighborhoods – PHDCN[4] and the Multi-Ethnic Study of Atherosclerosis – MESA.[1] We chose this methodology exactly because it permits the spatial delimitation of cutting edges. People who live in each local neighborhood will have the same SES and contextual characteristics as stated by Skater analysis based on the selected census variables. In addition, this methodology was chosen based on feasibility and our interest in comparing our results with other large international and national studies.

3. More details on identifying the neighbourhood principal components and clusters are needed in the methods. As both of these methodologies are very data-driven, a more thorough description of these methods, and an acknowledgement of how the results may differ widely in a different sample is needed.

This paragraph has been edited to make it clearer and more understandable.

4. The use of tertiles may be especially problematic when stratifying by SES cluster, as the difference in the means (based on supplementary table S1) are all quite small. The authors should be mindful of whether any statistical significance is indeed meaningfully different. The authors argue for the use of the tertiles as the three scales had different ranges and were not easily comparable to one another. However, the authors could have used other methods (such as standardization, or using defined cut-offs based on the literature) to create meaningful comparison groups.

We politely disagree with the comment that the differences in means are small in the sense that they are different enough to correspond to the bottom tertile, the middle tertile, and the highest tertile distribution of responses in our sample. As we mentioned in the first round of reviews, grouping scores into tertiles allow us to classify people into relative terms compared to the rest of the sample. While some previous studies that have assessed single neighborhood variables (i.e. either social cohesion or perceived violence) in relation to obesity have used such scores as continuous variables,[16,17] other studies that report on several neighborhood variables – like us – have also used tertiles to facilitate comparison of associations across neighborhood variables.[18] To the extent of our knowledge, "defined cut-off points based on the literature" do not exist, which is why other researchers before us have used the scores as continuous scores or, like us, have categorized them in tertiles of other cut-off points based on sample distributions (e.g. quartiles). We do not believe using other methods of categorization, such as standardization, would lead to qualitatively different results as those currently presented in the paper.

5. Building on the previous items, a more thoughtful discussion of the limitations of the research is needed. Any one of the previous comments would be considered a major limitation in a research study, and more transparency in what the study can (and cannot do) is needed.

---

[16] Cohen, D.A., Finch, B.K., Bower, A., Sastry, N. (2006) Collective efficacy and obesity: The potential influence of social factors on health. *Soc Sci Med* 62(3): 769-778.
[17] Guilcher, S.J.T., Kaufman-Shriqui, V., Hwang, J., et al. (2017) The association between social cohesion in the neighborhood and body mass index (BMI): An examination of gendered differences among urban-dwelling Canadians. *Prev Med.* 99: 293-298.
[18] Burdette, H.L., Wadden, T.A., Whitaker, R.C. (2006) Neighborhood safety, collective efficacy, and obesity in women with young children. *Obesity* 14(3): 518-525.

We have added the following information to the limitations' section based on the reviewer's prior criticisms (additions underlined):
Data collection was based on validated questionnaires and scales, as well as direct body measurements which allowed us to estimate obesity based on measured weight and height as opposed to self-reports. The neighborhood social environment variables, however, are all self-reported and we did not have access to objective measures of crime/violence in the neighborhood. Moreover, we aggregated individual-level scores from the social cohesion, perceived safety, and perceived violence scales to the neighborhood level so that all participants in the same neighborhood would have the same level of exposure. While this is standard procedure for the use of these scales,[28,29] the aggregate values are based only on the ELSA-Brasil sample and not on a representative sample of neighborhood residents.

6. Related to the previous item, there is concern that the authors made conclusions which were not justified by the results. In the discussion, the authors describe how the study results can be extrapolated to future interventions. This feels like a big leap to make based on the study that was conducted, and the authors should not overextend or extrapolate beyond what their study results demonstrate.

In the previous round of reviewers, this reviewer requested that we include "more tangible suggestions on what policy, researchers, or communities can do to try to help combat this problem together" in the Discussion section of the manuscript. We responded that we have suggested possible interventions at the neighborhood level that could address the social characteristics we found to be associated with obesity among Brazilian women (social cohesion and perceived violence). As to not stretch our findings, we also suggested in the conclusions that intervention researchers should test the proposed interventions addressing social cohesion and perceived violence. We are not saying these interventions will work but that our results suggest that they may work; however, they must be tested. We believe we are proposing tangible solutions, as previously requested by this reviewer, while being careful with our language to not overextend our findings.

Minor comments:
7. Introduction: The necessity of a sex-stratified analysis/objective of the paper is not well justified with the current introduction.

We have expanded the objective/hypothesis section of the Introduction to address this issue as follows (additions underlined):

To fill-in such gaps in the literature and taking advantage of a rich and georeferenced dataset based on six large cities in Brazil, the aim of this study was to investigate if the neighborhood social environment – including social cohesion, perceived safety and violence – was associated with obesity among Brazilian adults, and to identify if this association varied by gender. Previous studies have found that social neighborhood characteristics are associated with obesity[5-8,19,20] and that neighborhood environments affect women more than men;[23,24] therefore, we hypothesized that the neighborhood social environment would be associated with obesity among Brazilian adults, particularly among women. Furthermore, we hypothesized that lower neighborhood socioeconomic status (SES) could modify individuals' perceptions of their neighborhood environment and, thus, influence obesity-related behaviors. Therefore, we also assessed if the association between the neighborhood social environment and obesity varied by neighborhood SES.

8. Methods: As previously mentioned in the last round of comments (item 7), the authors create neighbourhoods based on methodology by Santos et al. However, an important difference between this paper and the methodology by Santos is the use of 'percent of white residents' in identifying neighbourhoods. We feel this is quite a departure in methodology from that by Santos et al., and should be justified.

As explained in the previous round of reviews, the study by Santos et al. (2010) was based on SES variables from the Brazilian Census from year 2000, while the current study is based on SES variables from the Brazilian Census 2010. In the 2010 Census, the question about education/schooling was not included, so the methodology was adapted and the variable about color/race (percentage of white *residents) was identified by principal component analysis as adequate to replace the education one. This additional information has been included in the revised version of the manuscript as follows (additions underlined):*

"Therefore, neighborhoods were constructed by combining contiguous census tracts with similar sociodemographic composition based on four variables from the Brazilian Census 2010:[25] number of people per household, proportion of children 0-4 years, mean income, and percent of white residents, following <u>an adaptation</u> of the methodology described by Santos et al. (2010).[24] <u>In their study, Santos et al. (2010)[24] utilized a spatial aggregation method based on SKATER (Spatial 'K'luster Analysis by Tree Edge Removal at TerraView software) to create clusters of contiguous census tracts based on the same sociodemographic characteristics listed above but with educational attainment instead of percent of white residents, as available in the Brazilian Census 2000.[24] The Brazilian Census 2010 did not include questions regarding education,[25] so percent of white residents was chosen as an adequate replacement variable based on principal component analysis."</u>

9. Table 1 should include p-values

As mentioned in the previous round of reviews, since we are following STROBE guidelines (as the journal requires), statistical comparisons of sample characteristics are not included:

STROBE guidelines (page 1643)[19]: "Inferential measures such as standard errors and confidence intervals should not be used to describe the variability of characteristics, and significance tests should be avoided in descriptive tables."

10. Regression tables should include the sample sizes
All tables include the sample sizes in the titles.

Reviewer: 3
Reviewer Name: Ruben Brondeel
Ghent University - Department of Movement and Sports Sciences
I appreciate the work that has been done by the authors. However, I feel some of the reviewers' remarks have been not been answered completely. For the first two comments, I acknowledge that these methods (categorizing dependent and independent variables) are common business in the literature, and that the other reviewers did not mention this point.

Response to response:
1. BMI categories 'normal weight', 'overweight' and 'obese' have been criticized by many authors for being arbitrary. The argument that differences between 22-23 are different from 24-25 has no real scientific basis. If the link with these categories needs to be made to help interpretation by clinicians, this could be done in a sub-analysis. Also, non-linearity can be handled using continuous variables.

We again politely disagree with the reviewer. We really do not think that studying obesity as an outcome needs to be further justified. Obesity, regardless of if we treat is as a medical condition on its own right or as a leading risk factor for other medical conditions such as heart disease, stroke, diabetes, hypertension, etc., is an important health outcome and a public health priority all around the world. Even though statistically speaking, using a continuous variable such as BMI would be preferable, our outcome of interest is obesity, not BMI. We are thankful for the reviewers' comments, which we truly believe have assisted us on improving the quality of this paper. However, we strongly believe that the authors should be the ones deciding on what their outcome of interest is. Moreover, even though the current BMI cut-off points have an arbitrary element, as argued by the reviewer, the reality is that these are the only currently acceptable international cut-off points we have, which were based on studies relating BMI with

---

[19] Vandenbroucke JP et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Med* 4(10; e297): 1629-1654.

mortality.[20,21] The only more-or-less accepted criticism to these cut-off points comes when the population under study includes individuals from the Asian continent or of Asian descent,[22] which is not our case.

With that said, we re-ran our analysis with BMI as the outcome, adjusting for age, education, skin color, ELSA site (as fixed-effect --- see also response to comment #7 below), and individual-level neighborhood scores, with main results displayed in Table R1 below. As seen in the Table, the results are qualitatively similar to those using obesity as the outcome, with the important caveat that in fully adjusted models (Model 2), only neighborhood perceived violence was significantly associated with BMI among women, but social capital was only marginally associated with BMI among women (p=0.0584). Still, we believe this shows our results are robust.

2. No justification for the tertiles can be found in comment 12 to reviewer 2. As in the comment about dichotomizing BMI, if specific contrasts need to be made - which doesn't seem to be the case in the article since the results are not interpreted in such a way - this can be done in a sub-analysis.

Please see response to reviewer 2, point #4 above.

3. There is no reason that an interaction effect would be more difficult to interpret. Interaction terms estimate the same differences as those in sub-group analyses, so it's just about representing the results in the correct way. Also, interaction terms would provide an insight in the differences in results between, for example, women and men.

We have included information regarding interactions in the paper, as highlighted below, but we have still chosen to display and discuss our gender-stratified results. Once interactions have been identified as significant, it is equally valid to present and interpret models with interactions, or present and interpret stratified models (while providing information on the interactions).

We have rephrased the statistical analysis section as follows:
"Hierarchical multilevel logistic regression models were ran as individuals (level 1) were nested within neighborhoods (level 2), and the outcome variable (obesity) was dichotomous. Model 1 included our independent variable of interest (social cohesion, perceived safety, or perceived violence) and age, while Model 2 was further adjusted by gender; education; skin color; ELSA sites; an interaction term between gender and social cohesion, perceived safety, or perceived violence; as well as individual-level scores on the social cohesion, perceived safety, and perceived violence scales for the models with neighborhood social cohesion, perceived safety, and perceived violence as predictors, respectively. This latter adjustment allowed us to account for individual variations in neighborhood perceptions and to obtain neighborhood effects above and beyond individual effects. Given that gender interactions for two out of our three independent variables of interest were significant (social cohesion interaction p-value=0.0077; perceived safety p-value=0.3569; perceived violence p-value=0.0363), we re-ran all models stratified by gender."

---

[20] National Institutes of Health. (1998) Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults-the evidence report. *Obes Res* 6(supplement 2): 51S–209S.
[21] WHO. *Report of a WHO Expert Committee. WHO Technical Report Series*. 854. Geneva, Switzerland: World Health Organization; 1995. Physical status: the use and interpretation of anthropometry.
[22] WHO Expert Consultation. (2004) Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 363: 157-163.

Table R1: Results from the multilevel linear regression model predicting body mass index by social cohesion, perceived neighborhood safety, and perceived neighborhood violence

	Model 1				**Model 2**			
	Women		**Men**		**Women**		**Men**	
	B (SE)	**p-value**	**B (SE)**	**p-value**	**B (SE)**	**p-value**	**B (SE)**	**p-value**
Neighborhood social cohesion								
Lowest tertile	**0.83 (0.20)**	**<.0001**	-0.15 (0.18)	0.4010	*0.40 (0.21)*	*0.0594*	-0.33 (0.20)	0.0959
Middle tertile	0.30 (0.16)	0.0703	-0.02 (0.14)	0.9018	0.15 (0.16)	0.3331	-0.05 (0.15)	0.7269
Highest tertile	Ref.		Ref.		Ref.		Ref.	
Neighborhood perceived safety								
Lowest tertile	0.30 (0.19)	0.1216	-0.06 (0.16)	0.7051	0.30 (0.20)	0.1247	0.13 (0.18)	0.4854
Middle tertile	-0.16 (0.17)	0.3397	**-0.30 (0.14)**	**0.0379**	-0.05 (0.16)	0.7428	-0.10 (0.15)	0.5013
Highest tertile	Ref.		Ref.		Ref.		Ref.	
Neighborhood perceived violence								
Lowest tertile	**1.00 (0.19)**	**<.0001**	-0.04 (0.16)	0.8063	**0.48 (0.21)**	**0.0230**	-0.01 (0.20)	0.9640

Middle tertile	0.09 (0.17)	0.6095	-0.24 (0.14)	0.0886	-0.02 (0.16)	0.9042	-0.19 (0.15)	0.2103
Highest tertile	Ref.		Ref.		Ref.			

Model 1 adjusted by age. Model 2 adjusted by age, education, skin color, ELSA site, and individual-level social cohesion, perceived safety, and perceived violence scores

4. My remark wasn't about the Pearson correlation. There is, in my opinion, a problem with the (face) validity of the items.
The reviewer's comment in the first round of reviews was: "Were there any ***correlations*** *between the various neighbourhood scores with one another? Is there any concern that they may be* ***multicollinea****r with one another and so are presenting 'redundant' information in some of the regression models?" (emphasis added). Given this comment, we ran correlations between the items and reported them back to the reviewer. We also included this information in the revised version of the manuscript.*

The face validity of the included items was not assessed by our team, nor were we able to find explicit reports of face validity – only other types of validity – from previous studies reporting psychometric characteristics of the included scales. However, given that face validity is subjective and dependent on for whom the items should be valued at face (to the respondents? To the authors? Other researchers?), in addition to having minimal influence on the objective content validity of a given scale,[23] we do not see this is a big issue in our study.

[23] DeVellis RF. *Scale Development: Theory and Applications. Second edition.* Applied Social Research Methods Series, Volume 26, March 2003.

5. (Some aspects of) social cohesion could be measured objectively. For example, with methods counting the number of contacts a person has during the day (e.g. smartphone Bluetooth data).

This is correct, though "number of contacts a person has" is often operationalized as a "social ties" construct,[4] which is a construct related to social cohesion but distinct from it.

6. PCA searches by definition for perpendicular dimensions. So clustering them in 3 categories and then naming them 'low', 'middle' and 'high', as if they are on 1 dimension, does not make sense. From this methodology, you would expect categories like 'low-low', 'low-high', 'high-low', 'high-high', in its most basic form.

Principal component analysis (PCA) is a multivariate statistical analysis technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated variables, which contain most of the original set information. A smaller set of uncorrelated variables is much easier to understand and use than a large set of correlated variables. Each principal component (PC) can then be seen as a new variable and used in further analysis. We reduced the four original census variables into two PCs that explained 87% of the variability of the data. With the PCA we identified which sets of variables explained most of the total variability, revealing what kind of relationship exists between them.

The new variables (PCs) were then used in a cluster analysis. The multivariate classificatory technique of clustering can be used when one wants to explore the similarities between areas by classifying them in groups (clusters), considering all the variables in each area simultaneously. The observations belonging to the same cluster are as similar as possible and always more similar to the elements of the same cluster than to elements of the other clusters. Therefore, the result of the cluster analysis is the classification of each area in one cluster. The interpretation of each cluster is empirical, made by a previous knowledge of the areas and by analyzing the scores of each component in the cluster. The result of a cluster analysis in one category for each cluster (low, medium, high in our study), rather than a classification based on surrounding areas, like in the local index of spatial autocorrelation (low-low, low-high, high-high, high-low).

7. I believe there is some misunderstanding in the use of a random vs fixed effect. The test presented in the comments is not relevant to this, since it tests a model with random effect to a model without any effect. Also, any variable (e.g. age groups) that explains some variance could give you a significant test result. The fact that the centers are very different, is no motivation for a random effect vs a fixed effect either. Both types of effects would adjust for the relatedness between participants in the same center. 'Centroa' could be introduced as both random or fixed term since it has only 6 categories and each category includes many participants. Fixed effects have the advantage that you get interpretable coefficients, and there are less assumptions in the calculation of the coefficients. A random effect really becomes necessary if there are so many categories (with usually relatively few participants), a fixed effect would no longer be interpretable.

Thank you for this clarification. All analyses were ran again removing ELSA sites as a random-effect variable and instead including it as a fixed-effect. The results do not qualitatively change, as shown below in Tables R2-R5. Tables and text have been updated accordingly in the new version of the manuscript.

Table R2: Results from the multilevel logistic regression model predicting obesity (BMI≥30 kg/m^2) by neighborhood social cohesion; gender-stratified (N= 6,092 women; 4,783 men)

	Model 1	Model 2
	OR (95%CI)	OR (95%CI)

	Women	Men	Women	Men
Neighborhood Social cohesion				
Lowest tertile	**1.43 (1.18-1.72)**	0.99 (0.81-1.21)	**1.25 (1.02-1.53)**	0.90 (0.72-1.13)
Middle tertile	1.14 (0.97-1.32)	0.96 (0.82-1.13)	1.07 (0.92-1.26)	0.95 (0.80-1.13)
Highest tertile	1.00	1.00	1.00	1.00
Age	**1.02 (1.01-1.03)**	1.00 (0.99-1.01)	**1.02 (1.01-1.03)**	1.00 (1.00-1.01)
Individual-level social cohesion			1.00 (0.98-1.02)	0.99 (0.97-1.01)
Education				
Primary or less			**1.46 (1.16-1.83)**	1.10 (0.87-1.40)
Secondary			**1.48 (1.28-1.70)**	1.10 (0.92-1.30)
University			1.00	1.00
Skin color				
Black			**1.86 (1.56-2.21)**	**1.45 (1.15-1.82)**
Brown			**1.38 (1.18-1.62)**	1.13 (0.95-1.36)
White			1.00	1.00
ELSA site				
Bahia			**0.62 (0.50-0.76)**	**0.56 (0.44-0.72)**
Espirito Santo			**0.70 (0.51-0.96)**	0.88 (0.63-1.23)
Minas Gerais			**0.75 (0.62-0.91)**	0.91 (0.74-1.11)
Rio de Janeiro			0.89 (0.71-1.11)	1.06 (0.84-1.33)
Rio Grande do Sul			0.91 (0.74-1.12)	0.91 (0.72-1.16)
Sao Paulo			1.00	1.00

Table R3: Results from the multilevel logistic regression model predicting obesity (BMI≥30 kg/m^2) by neighborhood perceived safety; gender-stratified (N= 6,092 women; 4,783 men)

	Model 1		Model 2	
	OR (95%CI)		OR (95%CI)	
	Women	Men	Women	Men
Neighborhood Perceived safety				
Lowest tertile	1.16 (0.97-1.38)	0.98 (0.82-1.18)	1.15 (0.95-1.39)	1.06 (0.86-1.30)
Middle tertile	0.94 (0.80-1.10)	0.92 (0.78-1.08)	0.96 (0.82-1.12)	0.98 (0.83-1.17)
Highest tertile	1.00	1.00	1.00	1.00
Age	**1.02 (1.01-1.03)**	1.00 (0.99-1.01)	**1.02 (1.01-1.03)**	1.00 (1.00-1.01)
Individual-level perceived safety			0.99 (0.97-1.02)	1.00 (0.98-1.03)
Education				
Primary or less			**1.48 (1.18-1.86)**	1.10 (0.87-1.39)
Secondary			**1.49 (1.30-1.71)**	1.09 (0.92-1.30)
University			1.00	1.00
Skin color				
Black			**1.87 (1.57-2.23)**	**1.44 (1.14-1.81)**
Brown			**1.39 (1.19-1.63)**	1.13 (0.95-1.36)
White			1.00	1.00
ELSA site				
Bahia			**0.59 (0.48-0.73)**	**0.56 (0.43-0.72)**
Espirito Santo			**0.66 (0.48-0.90)**	0.89 (0.64-1.24)
Minas Gerais			**0.71 (0.59-0.86)**	0.92 (0.75-1.12)
Rio de Janeiro			0.84 (0.67-1.05)	1.05 (0.83-1.32)
Rio Grande do Sul			0.87 (0.71-1.07)	0.92 (0.72-1.17)
Sao Paulo			1.00	1.00

Table R4: Results from the multilevel logistic regression model predicting obesity (BMI≥30 kg/m^2) by neighborhood perceived violence; gender-stratified (N= 6,092 women; 4,783 men)

	Model 1		Model 2	
	OR (95%CI)		OR (95%CI)	
	Women	Men	Women	Men
Neighborhood Perceived violence				
Lowest tertile	**1.51 (1.27-1.80)**	1.08 (0.90-1.30)	**1.28 (1.04-1.56)**	1.07 (0.86-1.34)
Middle tertile	1.07 (0.91-1.25)	0.98 (0.83-1.15)	1.03 (0.88-1.20)	0.99 (0.84-1.18)
Highest tertile	1.00	1.00	1.00	1.00
Age	**1.02 (1.01-1.03)**	1.00 (0.99-1.01)	**1.02 (1.01-1.03)**	1.00 (1.00-1.01)
Individual-level perceived violence			0.98 (0.96-1.00)	0.99 (0.96-1.02)
Education				
Primary or less			**1.42 (1.13-1.78)**	1.08 (0.86-1.38)
Secondary			**1.44 (1.25-1.66)**	1.08 (0.91-1.29)
University			1.00	1.00
Skin color				
Black			**1.82 (.153-2.17)**	**1.43 (1.13-1.81)**
Brown			**1.37 (1.17-1.61)**	1.13 (0.94-1.35)
White			1.00	1.00
ELSA site				
Bahia			**0.56 (0.45-0.70)**	**0.54 (0.42-0.70)**
Espirito Santo			**0.67 (0.49-0.92)**	0.89 (0.64-1.23)
Minas Gerais			**0.70 (0.58-0.84)**	0.92 (0.75-1.11)
Rio de Janeiro			0.81 (0.65-1.02)	1.02 (0.81-1.29)
Rio Grande do Sul			0.84 (0.68-1.03)	0.90 (0.71-1.15)
Sao Paulo			1.00	1.00

Table R5: Results from the multilevel logistic regression model predicting obesity (BMI≥30 kg/m^2) by neighborhood social cohesion, by perceived safety, and by perceived violence independently, stratified by neighborhood socioeconomic status (SES) and gender[1]

	High SES		**Intermediate SES**		**Low SES**	
	OR (95% CI)		**OR (95%CI)**		**OR (95%CI)**	
	Women	**Men**	**Women**	**Men**	**Women**	**Men**
Social cohesion						
N	2799	2144	1882	1371	1410	1268
Lowest tertile	**1.48 (1.10-1.99)**	1.03 (0.76-1.42)	0.86 (0.59-1.26)	0.95 (0.63-1.45)	1.43 (0.98-2.10)	0.92 (0.59-1.44)
Middle tertile	1.06 (0.82-1.37)	1.05 (0.80-1.38)	1.03 (0.77-1.37)	0.83 (0.59-1.16)	0.98 (0.73-1.33)	0.83 (0.58-1.20)
Highest tertile	1.00	1.00	1.00	1.00	1.00	1.00
Perceived safety						
N	2797	2144	1881	1371	1408	1268
Lowest tertile	1.09 (0.80-1.48)	1.01 (0.73-1.40)	0.81 (0.57-1.16)	1.11 (0.74-1.66)	1.38 (0.93-2.02)	1.19 (0.76-1.86)
Middle tertile	1.00 (0.77-1.30)	0.87 (0.66-1.14)	0.84 (0.63-1.13)	0.98 (0.70-1.37)	1.27 (0.94-1.71)	0.99 (0.69-1.43)
Highest tertile	1.00	1.00	1.00	1.00	1.00	1.00
Perceived violence						
N	2792	2134	1873	1369	1406	1267
Lowest tertile	1.00 (0.73-1.37)	1.21 (0.85-1.72)	1.22 (0.84-1.76)	1.03 (0.68-1.56)	**1.92 (1.28-2.90)**	1.02 (0.63-1.66)
Middle tertile	0.87 (0.67-1.12)	1.08 (0.83-1.40)	1.01 (0.75-1.38)	0.86 (0.60-1.22)	**1.70 (1.23-2.34)**	1.03 (0.72-1.49)
Highest tertile	1.00	1.00	1.00	1.00	1.00	1.00

[1]All models adjusted by age, education, skin color, ELSA site, as well as by individual-level social cohesion, perceived safety, and perceived violence scores for the neighborhood social cohesion, perceived safety, and perceived violence models, respectively.
[2]The perceived violence scale was constructed so that a higher score indicated a lower perceived violence. Therefore, the *lowest tertile* category represents neighborhoods with the highest perceived violence.

New comments:
Introduction:
8. The hypothesis are not completely based on the introduction. There are no arguments to include gender and neighborhood SES as moderators to the model, or why other variables are not considered as moderators (e.g. age).

Please see response to reviewer 2, point #7 above.

Discussion:
9. In the discussion, it becomes clear that food and physical activity neighborhood variables were available. These variables should have been included in the model, or it should be argued why they are not. At first sight, they seem to possibly confound the reported associations.

In the current study we were interested in the effects of the social neighborhood environment on obesity, not the physical neighborhood environment, which has already been studied in previous publications with ELSA-Brasil data.[24,25] Even though the physical neighborhood environment may affect obesity risk through similar proposed mechanisms than the social neighborhood environment (e.g. through its impact on physical activity levels), we do not think they would act as confounders in the associations between the social environment and obesity.

10. P20 line 25: it doesn't really matter that there are no gender differences in the social environment variables, the associations could still result from different mechanisms. (difference between means and covariances).

We have removed this problematic sentence from the discussion.

11. P20 line 46: as in previous comment, the first reason is not really relevant. The same or different perceptions could both result in similar or different associations. (see p22 line 3, for an example of this).

In this case, we are reporting what previous research has pointed out as possible reasons for women being more affected by their neighborhoods than men. As such, we are leaving this sentence as is.

12. P22 line 48: as mentioned by another reviewer, the causality is put too strong here. There should be at least a more thorough discussion on the exact mechanism that social cohesion influences obesity.

The Discussion section includes two paragraphs discussing what previous studies have found in relation to social cohesion and obesity and the hypothesized mechanisms linking these two; the latter is copied here for reading ease:

"Social cohesion is hypothesized to act as a buffer from neighborhood-related stress and, through this mechanism, be protective of obesity[5] Cohen et al. (2006) also suggest that adults in neighborhoods with higher social cohesion may be willing to intervene in aspects of the neighborhood that influence weight-related behaviors; for example, setting up sports leagues or influencing local food stores to carry healthier offerings.[5] However, the opposite can also be true, with residents in high-social-cohesion neighborhoods uniting for negative things as they pertain to obesity, for example, standing against soda taxation or against bans of unhealthy vending machines.[4]"

---

[24] Chor D, Cardoso LO, Nobre AA, et al. (2016) Association between perceived neighbourhood characteristics, physical activity and diet quality: results of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *BMC Public Health.* 16:751.
[25] Pereira de Castro OC, Arauji Nobre A, Ribeiro de Castro IR, Chor D, Harter Griep R, de Oliveira Cardoso L. Does context influence the Body Mass Index of Brazilian workers? Results from the ELSA-Brasil study baseline. Manuscript submitted for publication.

As for the sentence in question in p22 line 48 [of the original submission], we say that "Increasing access to safe public spaces may also help increase social cohesion and thus decrease obesity risk" [emphasis added]. We have now rephrased to "increasing access to safe public spaces may also help increase social cohesion and thus may decrease obesity risk."

Strengths and limitations:
13. P23 - line 20: Could you strengthen the argument by providing some comparison between the general population and the sample? Since poverty and poor neighborhoods is very significant for the reported associations, it would be good to have some more information on this (e.g. percentage living in low/middle/high SES neighborhoods).

There is additional information in the same limitations' paragraph, copied here for ease – please see underlined portions:

"This study is based on civil-servants in six large cities in Brazil; therefore, our sample excludes the extremely poor and unemployed and so our results may only be generalizable to Brazilian adults with stable employment. However, the ELSA-Brasil sample is diverse in terms of sociodemographic characteristics, including diverse regions within Brazil. Moreover, whereas the ELSA-Brasil sample has, on average, a higher income and social class than the residents of the six included cities,[25,44] the ELSA-Brasil sample has a similar prevalence of obesity and obesity-related behaviors (i.e. diet and physical activity patterns) than the Brazilian population at large.[45]"

14. P23 – line 27: it does not really matter that the prevalence of obesity is about the same in the sample as in the Brazilian population. It is somewhat encouraging, but it is possible that within two groups with the same prevalence of obesity, there are two totally different mechanisms that lead to obesity (see remarks above).
We agree with this comment, but see no obvious action steps attached to it. No changes were made based on this comment.

Minor remarks:
15. P6 line 17: … to work … to be protective
This change has been made.

16. P23 lines 44 – 51: it is great to refer to selective residential mobility. Is it possible to rewrite this argument to make this clearer to the reader.
The last sentence of the Limitations section has been rephrased as follows:
"While some researchers question the validity of associating neighborhood-level variables with health outcomes due to people self-selecting into neighborhoods,[46] the ELSA-Brasil population is highly stable, with an average length of residence in their current neighborhood of 15 years."

17. P24 line 18: replace 'living in poverty' by 'living in poor neighborhoods'.
This change has been made.

VERSION 3 – REVIEW

REVIEWER	Lisa Kakinami and Anna Smyrnova (co-review) Concordia University, Montreal Canada
REVIEW RETURNED	25-Jul-2019

GENERAL COMMENTS	We appreciate the authors' work on this manuscript. While many of our previous comments were addressed in the 'response to reviewers' some were not incorporated into the manuscript as

	acknowledged limitations (when appropriate). It is perhaps likely that future readers will have the same concerns or questions that we raised, thus we ask that the most pertinent items also be included in the manuscript (as stated below). We can sympathize that the authors may be frustrated with the review process and respectfully would like to point out that all comments are made in the spirit of constructive criticism. Comment 1: We are well aware of the utility of self-perceptions of the neighborhood and were not arguing against their use in the study. The authors' argument for the aggregated self-reported perceptions (for shared context) and the individual self-reported perceptions (for individual differences) is well received and pertinent to the manuscript. We ask that the authors include snippets of their argument from their 'response to reviewers' in the discussion. References to these self-reported perception scales being "designed to be analyzed aggregated as contextual attributes" would be necessary to include in the manuscript as well. Comment 2: While it is undoubtedly true that previous studies have categorized people within each polygon (regardless of if they were on the edge or the center) to be the 'same', we feel that the fact that previous studies also did this is not a strong enough argument for continuing to make this analytic decision. We feel this analytic decision introduces error (or worse – bias), and should at least be acknowledged as a limitation in the discussion. Comment 4: The authors use tertiles calculated based on selected neighborhoods (resulting in the distribution of responses within the sample only). While we can appreciate the utility of tertiles in research, we politely ask authors to consider 'all' options before resorting to such a data-driven approach. In their 'response to reviewers' the authors argue that the use of tertiles is meaningful as the means between groups is "different enough to correspond to the bottom tertile, the middle tertile, and the highest tertile distributions of responses". While we still have conceptual concerns of whether or not the bottom/middle/highest tertiles are still meaningfully distinctive from a population-health standpoint, we see that the authors will not likely budge on this. Since the use of tertiles is a bit of a philosophical point of contention in the scientific community, we respect their decision and right to use tertiles. However, we would still like to point out one potential issue, which is getting to the true crux of our concern with tertiles. In Table 1, we note that for the 'social cohesion' variable, the lowest, middle, and highest tertiiles are approximately 19%, 50%, and 32% of the sample, respectively. The large discrepancy in proportions between the lowest and middle tertiles (19% and 50%) amounts to 3500 people who are 'right on the boundary' between lowest and middle tertiles. In other words, as the range for the lowest tertile is 5-16.3, and the range for the middle tertile is 16.3-18, this suggests that 3500 people have a social cohesion value of 16.3. Are the authors not concerned that these people who are right on the boundary between two tertiles may be misclassified? This

	same issue can be seen for the other tertile variables. Unless we are making an error in our understanding of the tables, we feel that in light of the fact that approximately 25% of the sample may be right on the boundaries, the authors should at least acknowledge these issues (misclassification, and data-dependent) as potential limitations of the use of tertiles in their discussion. Comment 6. We appreciate that the authors previously expanded their discussion section to address our previous concern of research/policy implications. Our previous comment 6 pertained specifically to the conclusions. While we feel that the discussion section is tempered in tone and does not overstate findings, the final paragraph of the manuscript (conclusions) is a bit too abbreviated and this tempering of language is lost. We suggest that as a way to not overstate conclusions, examples of the proposed interventions or future research investigating social cohesion and perceived violence (eg: further research into neighborhood watches, increased greenery, access to public spaces, etc) within that concluding paragraph will help ground the reader into what the authors rightfully feel are feasible future research and changes that can be made. New comments: -What does it mean that only the aggregated tertiles of these neighborhood variables were significant (and only among women) for obesity risk, but not the individual-level perceptions of these neighborhood variables? Quite honestly we are unclear of how to interpret this and would appreciate the authors' insight and thoughts into this in their discussion. Do the authors interpret these results to mean that it's the "collective" perception that is associated with obesity, and not any individual's perception? This logic is a bit hard to follow, and a thoughtful paragraph on the meaning of these aggregated vs individual differences would be important to include. -In response to Reviewer 3, the authors state diet and physical activity are not likely confounders between the social environment (as measured by the authors) and obesity risk. We respectfully disagree that these would not be confounders, and their omission as covariates should be acknowledged as a limitation in the discussion. -The Cronbach's alphas of the three scales are not great (0.60, 0.67, 0.71) and might suggest that future studies should be using scales that have better internal consistency (for the discussion section). -Typo on page 9: 'The outcome of this study…' should be BMI >= 30 (is currently BMI >30)

VERSION 3 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:
Reviewer: 2

Reviewer Name: Lisa Kakinami and Anna Smyrnova (co-review)
Institution and Country: Concordia University, Montreal Canada

We appreciate the authors' work on this manuscript. While many of our previous comments were addressed in the 'response to reviewers' some were not incorporated into the manuscript as acknowledged limitations (when appropriate). It is perhaps likely that future readers will have the same concerns or questions that we raised, thus we ask that the most pertinent items also be included in the manuscript (as stated below). We can sympathize that the authors may be frustrated with the review process and respectfully would like to point out that all comments are made in the spirit of constructive criticism.

Comment 1: We are well aware of the utility of self-perceptions of the neighborhood and were not arguing against their use in the study. The authors' argument for the aggregated self-reported perceptions (for shared context) and the individual self-reported perceptions (for individual differences) is well received and pertinent to the manuscript. We ask that the authors include snippets of their argument from their 'response to reviewers' in the discussion. References to these self-reported perception scales being "designed to be analyzed aggregated as contextual attributes" would be necessary to include in the manuscript as well.

We have incorporated in the Methods section – when discussing that the individual-level social cohesion, perceived safety, and perceived violence scores were aggregated at the neighborhood level – that we did this because the scales were designed to be aggregated as contextual characteristics (1st paragraph of page 11 in the marked copy of the manuscript).

Comment 2: While it is undoubtedly true that previous studies have categorized people within each polygon (regardless of if they were on the edge or the center) to be the 'same', we feel that the fact that previous studies also did this is not a strong enough argument for continuing to make this analytic decision. We feel this analytic decision introduces error (or worse – bias), and should at least be acknowledged as a limitation in the discussion.

This has been included as a limitation.

Comment 4: The authors use tertiles calculated based on selected neighborhoods (resulting in the distribution of responses within the sample only). While we can appreciate the utility of tertiles in research, we politely ask authors to consider 'all' options before resorting to such a data-driven approach.

In their 'response to reviewers' the authors argue that the use of tertiles is meaningful as the means between groups is "different enough to correspond to the bottom tertile, the middle tertile, and the highest tertile distributions of responses". While we still have conceptual concerns of whether or not the bottom/middle/highest tertiles are still meaningfully distinctive from a population-health standpoint, we see that the authors will not likely budge on this. Since the use of tertiles is a bit of a philosophical point of contention in the scientific community, we respect their decision and right to use tertiles.

However, we would still like to point out one potential issue, which is getting to the true crux of our concern with tertiles. In Table 1, we note that for the 'social cohesion' variable, the lowest, middle, and highest tertiiles are approximately 19%, 50%, and 32% of the sample, respectively. The large discrepancy in proportions between the lowest and middle tertiles (19% and 50%) amounts to 3500 people who are 'right on the boundary' between lowest and middle tertiles. In other words, as the range for the lowest tertile is 5-16.3, and the range for the middle tertile is 16.3-18, this suggests that 3500 people have a social cohesion value of 16.3. Are the authors not concerned that these people who are right on the boundary between two tertiles may be misclassified? This same issue can be seen for the other tertile variables. Unless we are making an error in our understanding of the tables, we feel that in light of the fact that approximately 25% of the sample may be right on the boundaries, the authors should at least acknowledge these issues (misclassification, and data-dependent) as potential limitations of the use of tertiles in their discussion.

Table 1 has been corrected to note that for social cohesion, for example, the tertile ranges go from 5 to <16.3 in the lowest tertile, 16.3 to <18 in the middle tertile, and 18 to 25 in the highest tertile; we apologize for this oversight. Social cohesion=16.3 is not included in both the low and the middle categories as the previous version of the Table may have suggested, only in the middle one, and only 20 individuals have this exact value. We are not sure how the reviewers came up with 3500 individuals belonging to that category, but this is not the case. Being neighborhood variables, the tertile cut-off points for social cohesion, perceived safety and perceived violence were estimated based on the number of neighborhoods (N=1902), not the number of individuals in the sample (N=11,456). So ~33% of the neighborhoods fall into the lowest social cohesion category, for example, but only 18.8% of the sample live in these neighborhoods.

We have included the use of tertiles as a potential limitation in the Limitations section.

Comment 6. We appreciate that the authors previously expanded their discussion section to address our previous concern of research/policy implications. Our previous comment 6 pertained specifically to the conclusions. While we feel that the discussion section is tempered in tone and does not overstate findings, the final paragraph of the manuscript (conclusions) is a bit too abbreviated and this tempering of language is lost. We suggest that as a way to not overstate conclusions, examples of the proposed interventions or future research investigating social cohesion and perceived violence (eg: further research into neighborhood watches, increased greenery, access to public spaces, etc) within that concluding paragraph will help ground the reader into what the authors rightfully feel are feasible future research and changes that can be made.

This has been incorporated in the Conclusions.

New comments:
-What does it mean that only the aggregated tertiles of these neighborhood variables were significant (and only among women) for obesity risk, but not the individual-level perceptions of these neighborhood variables? Quite honestly we are unclear of how to interpret this and would appreciate the authors' insight and thoughts into this in their discussion. Do the authors interpret these results to mean that it's the "collective" perception that is associated with obesity, and not any individual's perception? This logic is a bit hard to follow, and a thoughtful paragraph on the meaning of these aggregated vs individual differences would be important to include.

As explained in the "Statistical analysis" section of the manuscript, the purpose of adjusting for individual-level scores is to account for individual variations in neighborhood perceptions and to obtain neighborhood effects above and beyond individual effects. For example, I may think that my neighborhood is very dangerous, while my neighbor may perceive that our neighborhood is very safe. The average between my neighbor and I would be that the neighborhood is neither dangerous nor safe. So, if we follow the analyses in the paper, we are using the aggregate values as we consider these to be the “true” neighborhood characteristics. However, we still need to adjust for individual variations as some responses within this aggregate may be disparate.

As the purpose of this adjustment is already explained in the Statistical analysis and we believe we are following standard procedures, we have decided against explaining this any further in other sections of the manuscript.

-In response to Reviewer 3, the authors state diet and physical activity are not likely confounders between the social environment (as measured by the authors) and obesity risk. We respectfully disagree that these would not be confounders, and their omission as covariates should be acknowledged as a limitation in the discussion.
We did not argue that diet and physical activity were not confounders in the association between the social environment and obesity; we argued that the neighborhood food environment and the neighborhood built environment (i.e. places in the neighborhood where one could potentially be physically active) are likely not confounders of the association between the social environment and obesity. Individual-level physical activity, in particular, is likely a mediator between the social environment and obesity – not a confounder – as we have mentioned in the article in several occasions.

-The Cronbach’s alphas of the three scales are not great (0.60, 0.67, 0.71) and might suggest that future studies should be using scales that have better internal consistency (for the discussion section).
This information has been included in the limitations section, including a citation which discusses the use and validity of these scales in the Brazilian context (which is beyond the scope of the current article).

-Typo on page 9: ‘The outcome of this study…’ should be BMI >= 30 (is currently BMI >30)
This has been corrected, thank you for catching this typo.