

## 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

<b>TITLE (PROVISIONAL)</b>	Change in drink purchases in 16 Australian recreation centres following a sugar-sweetened beverage reduction initiative: an observational study
<b>AUTHORS</b>	Boelsen-Robinson, Tara; Orellana, Liliana; Backholer, Kathryn; Kurzeme, Ariana; Jerebine, Alethea; Gilham, Beth; Chung, A; Peeters, A

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Carlos Caro Health Policy and Management, University of North Carolina.
<b>REVIEW RETURNED</b>	20-Mar-2019

<b>GENERAL COMMENTS</b>	<p>I appreciate the opportunity to review this manuscript entitled: "Change in drink purchases in 16 Australian recreation centers following a sugar-sweetened beverage reduction initiative: an observational study". Following the experience of many countries to regulate and reduce the exposure to SSBs (e.g. Chile), this study documents the effects of similar strategies at the community level. Below are my major comments, with no particular order:</p> <ol style="list-style-type: none"> <li>1. The meta-analysis method is not defined in the manuscript. Key assumptions, beyond weighting, should be explicit. Similarly, the ITSA method should be explained in detail in the supplement file and briefly in the manuscript. Key assumptions have been omitted and interpretation relies heavily on the particular approach used (saying "usual ITSA assumptions" is not enough). Making a reference to an user-written Stata command is not an appropriate way to discuss the methods used, since specification varies with the dependent variable and research question at hand. A good example of describing methods can be found below:  Hsu, J. C., Wei, C. F., &amp; Yang, S. C. (2019). Effects of removing reimbursement restrictions on targeted therapy accessibility for non-small cell lung cancer treatment in Taiwan: an interrupted time series study. <i>BMJ Open</i>, 9(3), e022293.</li> <li>2. References are not correct in the supplement file and Manuscript. E.g. In the methods (S2), first reference is 26 where they are indeed citing 24.</li> <li>3. What do authors refer as "monthly attendance"? I assume is the number of visitors to each center but it should be explicit. In addition, the approach of "combining" months where some centers were closed violates the basic model assumptions (key issue).</li> </ol>
-------------------------	--

	<p>4. Authors should present graphs showing the change in trend and level before and after treatment, depending on how they define the post-period. There is no reason apriori to assume that the main effects should be measured at the end of the implementation period, unless proven otherwise.</p> <p>5. Given how beverages are classified under the intervention, we expect correlated errors due to substitution across categories that not necessarily are driven by the intervention. This point should be incorporated in the analysis (i.e. estimate all time series jointly and allow free error correlation) or at least be explicit about the assumptions and limitations.</p> <p>6. Given the way that the intervention was implemented, why not take an event-study approach? Using ITSA authors are obscuring the effects that are given by variation on the implementation strategies across centers (e.g. exclusion of diet carbonated drinks).</p> <p>7. Meta-analysis figures should all have the same order of centers in order to compare across different beverage categories. Moreover, is there any explanation of the substitution patterns observed within centers? this is key and largely ignored in the manuscript. Part of the success on implementation is not the reduction of "red" drinks but rather substitution towards healthier drinks, since consumer might easily substitute "red" drinks outside centers (e.g. at home). This should be discussed in detail.</p> <p>8. Inter-rater reliability seems low. Authors should explain why.</p> <p>9. For the reduction on drink varieties, baseline number should be provided as reference (or % change).</p> <p>10. Following a previous point, why stratification by initiative implementation was not conducted? it seems critical to discuss results of the initiative.</p> <p>11. Why sales value matter? What do authors expect to infer from comparing change in volume versus monetary value? Similarly, when discussing change in the volume of sales (relative to total sales) there is a key problem with implicit correlation across groups being ignored in the estimation. Also, how were the CI calculated in this case?</p> <p>12. Saying that weighting does not affect the outcomes of meta-analysis is insufficient, specially when methods nor output is available on the supplement file.</p> <p>13. Authors should be extremely careful with language. Sentences such as "ITSA is the strongest method for the evaluation of natural experiments" is extremely unfortunate and misguided.</p> <p>14. How does ITSA mitigate the absence of a "control" center? it seems that authors are unclear about the assumptions underlying the method.</p> <p>15. Results suggest a decline of red drinks without further substitution. A natural approach should be a qualitative follow-up to understand whether this represent a total reduction on SSBs intake or substitution at home. As is, this study cannot argue to</p>
--	---

	recommend this strategies as the claim above cannot be confirmed nor disprove (e.g. assuming that costumers are switching to water fountains is speculative). Overall, the conclusion is not supported by the results provided.
--	---

<b>REVIEWER</b>	Penny Breeze University of Sheffield, United Kingdom
<b>REVIEW RETURNED</b>	09-Apr-2019

<b>GENERAL COMMENTS</b>	<p>Main limitations</p> <ol style="list-style-type: none"> <li>1. The statistical methods are not sufficiently described. This section should be re-written so that it is clear exactly what statistical models have been adopted.</li> <li>2. The aims of the study and results are not consistent.</li> </ol> <p>All comments</p> <p>Introduction</p> <ol style="list-style-type: none"> <li>1. The introduction presents some background to the research in this area but excludes other very similar ITSA studies of sugar drinks reduction initiatives published. I was involved in similar study looking at price increases alongside promotion and product position changes in leisure centres in the UK also uses sales data and ITSA methods.</li> </ol> <p>Method</p> <ol style="list-style-type: none"> <li>1. More information is needed to describe the type of statistical model used. It is not sufficient to refer to the stata command used. Ideally the methods should be described in more detail within the main manuscript, with additional detail provided in the appendix. More detail is needed on the methods used for the meta-analysis and a justification given for equal weighting.</li> <li>2. More detail needs to be provided on the choice of methods for this analysis. Meta-analysis is not the only method available for allowing heterogeneity between study site so it needs justification.</li> <li>3. I agree with adjustment for seasonal changes in sales for the analysis. However, grouping months by season may not be the most appropriate way of defining these changes. For example one might expect that months with school holidays may see peaks in attendance independent of weather related changes. It would be useful to provide some descriptive statistics of sales data to justify your approach.</li> <li>4. It would be useful to see some illustration of the magnitude of sales, fluctuations over time and how they change during the pre, implementation and post-implementation period.</li> <li>5. It is not clear to me how the timing of the initiative work. Firstly, did all centres start the initiative at the same time or does the post-initiative period vary by centre? Secondly, it is not clear why the period Dec 2015-Dec 2016 was excluded from the primary analysis. This is particularly important in looking at how people adapt to the changes in drinks provision over time. I can see the benefit of looking at the two endpoints separately, but would prefer to see both within the main text.</li> <li>6. It is not conventional to reference results from the supplementary material in the methods section. If sensitivity analyses are performed they should be described in full within the methods or supplementary material.</li> </ol> <p>Results</p> <ol style="list-style-type: none"> <li>1. Given the objective of the study I am surprised to see the first section of the results address whether the initiative was successfully adopted by the centres. It may be necessary to</li> </ol>
-------------------------	--

	<p>rethink what is reported in the results section or re-word the aims of the study.</p> <p>2. It is not clear how the availability of drinks was assessed. I was under the impression that the study had access to sales data. Is it possible that there is wastage at the sites if products are not sold? Wouldn't this affect the accuracy of the availability analysis?</p> <p>3. The statement of inter-rater reliability is not suitably placed and requires further comment to make it clear what this refers to. Does this mean that there was some variation between centres as to how products were categories into the red-amber-green groups? If so the level of disagreement is surprising. Can you identify which types of products were most likely to be mis-categorised? From the methods the products were categories by a dietitian, was this done by the research team or locally by the centres to implement the policy?</p> <p>4. Given that the centres did not introduce the policy by Dec2015 I am even more surprised that the data you have beyond this period has not been used in the primary analysis.</p> <p>5. I think the results section would benefit from a table or figure illustrating the variation in results between the Dec 2015 and Dec 2016 end-points.</p> <p>6. "Stratification by variables of interest" is ambiguous. What are the variables of interest?</p> <p>Discussion</p> <p>1. I think the discussion of how the results compare with other studies could be improved. The manuscript refers to an Australian and Canadian study but provides no information on whether the policies implemented were similar to this one.</p> <p>2. It is interesting to note that a similar study in the UK found that a SSB initiative decreased sales of SSB, and increased sales of diet carbonated drinks, but did not observe changes in healthy options (fruit drinks or water).</p>
--	---

<b>REVIEWER</b>	Cordia Chu Griffith University, Australia
<b>REVIEW RETURNED</b>	07-Jun-2019

<b>GENERAL COMMENTS</b>	It would be helpful for the authors to explain clearly the intervention strategies applied by the 16 Australian recreation centres for the sugar-sweetened beverage reduction initiative.
-------------------------	---

<b>REVIEWER</b>	Beatrice Biondi University of Bologna, Italy
<b>REVIEW RETURNED</b>	23-Jul-2019

<b>GENERAL COMMENTS</b>	<p>As a general comment, the paper looks promising; data and study settings can form the base for a valuable and sound research. However, the application of the methodology and the displaying of results is not straightforward. From my point of view – which is the one of a statistician working with quasi-experimental methods, but not in the field of public health nor of clinical studies – the paper lacks in explanations and justifications for the researchers' choices of the statistical analysis, which affect the reliability of final results. I provide below some comments, with specific reference to the implementation and presentation of statistical methods in the paper, which I believe can improve the research overall. Points 1-5 represent the main weaknesses of the study, the other points are minor revisions in text.</p>
-------------------------	--

	<p>1. Definition of the intervention: after an in-depth reading, I understood that the original aim of what you called a reduction initiative was instead a complete removal of “red” SSB, and recreation centres applied it to a different degree (only one removed all red SSB). At page 4, lines 15-20, you explain the aim of the initiative, providing a very clear table for classification of beverages. Here, you should also clarify that not all centres applied this to the same extent. This is extremely important also for the assumptions behind the model: this means that the intervention is somehow different for all centres. You must provide descriptive statistics on the share of red SSB reduction implemented by each centre. And it is essential to know whether and when the change in assortment occurs: in order to estimate the effect of the reduction, assortment should (I) remain nearly stable in the pre-intervention period, (II) change with the intervention, (III) remain again nearly stable in the post-intervention period.</p> <p>2. Gradual implementation: the removal of red SSB was implemented gradually during the implementation year, at least this is what I understood from the text. This could create problems in the estimation of ITS. Indeed, ITS “is not appropriate when (...) the intervention is introduced gradually or at more than one time point” (Kontopantelis et al., 2015), you should control for this issue and provide evidence of how reduction share changes at the beginning and at the end of the implementation period.</p> <p>3. Period of implementation: How you considered pre- and post- intervention periods is not at all clear in the paper. In the ITS procedure you should identify a cut-off time point to divide the observations in pre and post, and it is confusing to me whether you take Dec 2014 or Dec 2015. At page 4, lines 5-13 you say that the intervention is one action of a series of actions, belonging to a broader policy, implemented over a three years period. Which period? Then you say that the intervention has been implemented over one-year period, completed by December 2015. Fig 1 then is divided in three periods, and the reader is unsure about the pre/post intervention time span. Finally, going to the methods section, you clearly say you used three periods. But what does change between implementation (Dec 2015) and post-implementation (Dec 2016)? Did the gradual removal that happened along 2015 reach a final point (share of red SSB) and remain constant in 2016? I would carefully reconsider the choice of choosing three periods, since it means two shocks took place: for instance, in their example Linden and Arbor (2015) consider the cases in which an intervention is introduced, withdrawn, and reintroduced; or an intervention is followed by a separate intervention at a later point in time. It seems to me that in your case only one intervention was considered. This could also explain some anomalous results you obtain (see point 14).</p> <p>4. Consideration of “amber” products: on page 5, lines 3-7 is said that amber products are not considered because some centres removed them and others reduced it. I think this is not a reason to not consider this type of products. In centres that implement a reduction, you could consider this type of products in the same way of red products. At page 9, line 53 you say the analysis is precluded. Instead, you could run an analysis on the 7 centres that sold amber products, and analyze purchase patterns focusing on substitution.</p> <p>5. ITSA specification: Linden and Arbor (2015): “when the treated group’s outcomes can also be contrasted with those of one or more comparison groups, the internal validity is further</p>
--	---

	<p>enhanced by allowing the researcher to potentially control for confounding omitted variables.” and Linden (2018) “When ITSA is implemented without a comparison group, the internal validity may be quite poor. Therefore, adding a comparable control group to serve as the counterfactual is always preferred.” Given these, I was surprised to see that you did not use a control group. Especially given that only 8 out of 68 centres have not accessible sales data (I assume the other 60 have accessible data), and you explicit that 5 centres did not commit to remove SSB. They could potentially be a perfect control group. At page 9, line 27 you say that control centres are unavailable, why is that? I encourage you in performing the analysis using the control group. You could use simple multiple group ITSA (Linden, 2015), or use a comparable control group, e.g. by matching on covariates related to centres (see Linden, 2018). You could consider also other methods, such as difference-in-difference. In any case, I believe the study results would greatly gather in reliability if you use a control group in the estimation.</p> <p>6. You did not explain how your “novel meta-analysis” is performed. I assume you averaged across centres. Did you explore the possibility of running a panel fixed-effect model, instead of running 16 different models? This might represent a more “elegant” solution, which allows to consider the difference among centres.</p> <p>7. To help readers, you could provide a table presenting, for each centre, the share of reduction of red SSB, the share of reduction of amber SSB, and the relative increase of green products after the implementation.</p> <p>8. In the Data Collection Section on page 4, you did not refer to Fig 2, which is only mentioned in the Results Section. I suggest you refer to Fig 2 while you explain the data collection, referring to the whole process displayed in it, since they do not match right now (i.e. exclusion criteria). On Page 4 Line 32: “Provided food and/or drink for sale from a kiosk or café” and canteens? I thought you were only interested in drink sales. I would rephrase this.</p> <p>9. Fig.2 what does “proposed model did not fit the time series” mean?</p> <p>10. MODEL: Linden and Arbor (2015) at page 481 display the formula of OLS regression used by ITSA. I think that this should be included Methods S2 section, to give the reader a clearer idea of the model, including an explanation of specific variables used in your study.</p> <p>11. Some questions follow from the reading of Linden and Arbor (2015) paper on ITSA. In particular, information is missing in your paper regarding tests:</p> <ul style="list-style-type: none"> <li>• Did you use newey or prais method in itsa estimation?</li> <li>• Did you test for the presence of autocorrelated errors?</li> <li>• Did you run robustness tests?</li> <li>• You could include in the paper the table of the beta estimates, with std error (not only beta and CI bounds).</li> </ul> <p>12. Price changes are not controlled for in the model. Did you check for changes in prices in the period considered? This could greatly affect your results.</p> <p>13. Results section: A graph of the trend in volume sales for red SSB on the period considered, cumulative for all the centres or even with a line for each centre, would help the reader in understanding the overall change in consumption over time.</p> <p>14. Page 7, line 38 here you say S6 table but you refer to S7; line 42 if I look to S5 table I see a 57.2 increase in green sales. How do you explain this increase after a 0 increase in 2015? At</p>
--	---



	<p>line 43 again you should refer to S7 table, where I see some differences among strata, indeed. Reformulate this paragraph according to the results shown in tables.</p> <p>15. Discussion Section: Page 9 lines 18-21: why do you think is best to avoid having all centres in one model? Please, motivate this sentence.</p> <p>References:  Kontopantelis, E. et al. (2015) 'Regression based quasi-experimental approach when randomisation is not an option: Interrupted time series analysis', <i>BMJ (Online)</i>, 350, pp. 1–4. doi: 10.1136/bmj.h2750.  Linden, A. (2018) 'A matching framework to improve causal inference in interrupted time-series analysis', <i>Journal of Evaluation in Clinical Practice</i>, 24(2), pp. 408–415. doi: 10.1111/jep.12874.  Linden, A. and Arbor, A. (2015) 'Conducting interrupted time-series analysis for single- and multiple-group comparisons', <i>Stata Journal</i>, 15(2), pp. 480–500. doi: 10.1177/1536867X1501500208.</p>
--	---

### VERSION 1 – AUTHOR RESPONSE

Reviewer comments	Author responses
<p>Reviewer: 1  Reviewer Name: Carlos Caro  I appreciate the opportunity to review this manuscript entitled: "Change in drink purchases in 16 Australian recreation centers following a sugar-sweetened beverage reduction initiative: an observational study". Following the experience of many countries to regulate and reduce the exposure to SSBs (e.g. Chile), this study documents the effects of similar strategies at the community level. Below are my major comments, with no particular order:</p> <p>1. The meta-analysis method is not defined in the manuscript. Key assumptions, beyond weighting, should be explicit. Similarly, the ITSA method should be explained in detail in the supplement file and briefly in the manuscript. Key assumptions have been omitted and interpretation relies heavily on the particular approach used (saying "usual ITSA assumptions" is not enough). Making a reference to an user-written Stata command is not</p>	<p>We thank the reviewer for this comment and the helpful suggestion of the Hsu et al paper. Based on yours and the other reviewers' recommendations, we have updated the Statistical Analysis section moving information from the supplement file and expanding to provide a description of the analytical approach rationale and the set of assumptions underlying each method (page 5-6).</p>

<p>an appropriate way to discuss the methods used, since specification varies with the dependent variable and research question at hand. A good example of describing methods can be found below: Hsu, J. C., Wei, C. F., &amp; Yang, S. C. (2019). Effects of removing reimbursement restrictions on targeted therapy accessibility for non-small cell lung cancer treatment in Taiwan: an interrupted time series study. <i>BMJ Open</i>, 9(3), e022293.</p>				
<p>2. References are not correct in the supplement file and Manuscript. E.g. In the methods (S2), first reference is 26 where they are indeed citing 24.</p>	<p>Thank you for this. We have moved the methods in the supplement to the main file so this is no longer an issue.</p>			
<p>3. What do authors refer as "monthly attendance"? I assume is the number of visitors to each center but it should be explicit. In addition, the approach of "combining" months where some centers were closed violates the basic model assumptions (key issue).</p>	<p>Monthly attendance represents the number of visitors to each centre per month, measured by scanned membership cards and manual purchases of once off visits. We have clarified this in the main manuscript on page 4 and page 5.</p> <p>The decision of "combining" Dec and Jan data for centres (3, 4 and 9) was based on the fact that Centre 9 was closed during 4 weeks over Christmas and Centre 3 and 4 were small centers with stadiums and no pools, which had very low attendance numbers over the December and January months. We considered that information in sales and attendance from Dec and Jan could be combined to represent "approximately one month". The change in the model parameterization to account for the "11 month" year has been added to the Statistical Analysis section (page 7).</p> <p>To assess the robustness of our conclusions and the influence of these three centres on the average results, we ran a new meta-analysis excluding these centres. Results for 'red' and 'green' packaged drinks volume and for package drink sales (Dec 2015) are shown in the table below for the reviewer. The main conclusions of the study remained unchanged when these centres were excluded. These results have not been included in the manuscript, however we are able to on request of the reviewer or editor.</p> <table><tr><td>Outcome</td><td>Change in sales compared to counterfactual, all centres.</td><td>Change in sales compared to counterfactual, excluding Centres 3,4</td></tr></table>	Outcome	Change in sales compared to counterfactual, all centres.	Change in sales compared to counterfactual, excluding Centres 3,4
Outcome	Change in sales compared to counterfactual, all centres.	Change in sales compared to counterfactual, excluding Centres 3,4		



		December 2015, (CI), %	& 9, December 2015, (CI), %
	Packaged 'red' volume	-46.2 (-53.2, - 39.1)	-48.3 (-56.1, -40.5)
	Packaged 'green' volume	-0.0 (-13.3, 13.2)	-4.7 (-17.6, 8.2)
	Packaged drinks sales value	-24.3 (-32.0, - 16.6)	-24.5 (-33.0, -16.0)
<p><b>4.</b> Authors should present graphs showing the change in trend and level before and after treatment, depending on how they define the post-period. There is no reason apriori to assume that the main effects should be measured at the end of the implementation period, unless proven otherwise.</p>	<p>Thank you for this suggestion. We have now included Figure S1 in the Supplement that illustrate the raw data, the predicted outcomes under the intervention and the predicted counterfactual outcomes for 2 centres. The 2 centres represented in the Figure S1 were selected to represent centres with different facilities (i.e. pools, stadiums) and selling different amounts of drinks (i.e. high or low).</p> <p>The time periods were based on YMCA's implementation aims as illustrated in Figure 1. Centres were given from the start of December 2014 until the end of November 2015 to implement the SSB reduction initiative. We considered three time periods as we were aware that the rate of implementation was likely to differ across the centres. Further, it was likely that some centres had not in fact been able to fully implement, or maintain, the SSB reduction initiative. The third time period started in December 2015 the month by which the initiative was intended to be fully implemented.</p> <p>We have now clarified in the Statistical Analysis section that we conducted an intention to treat analysis, i.e. we assessed the effect of an initiative that was a directive of the central organisation, i.e. the head office of YMCA, but that was implemented on the ground by the manager and staff at each centre. Therefore it was important for the organization to have evidence of the effect of the initiative by the time the initiative was supposed to be fully implemented (Dec 2015) We further chose to assess the initiative outcomes at December 2016 to gain insight into the sustained impact, one year following expected implementation.</p>		
<p><b>5.</b> Given how beverages are classified under the intervention, we expect correlated errors due to substitution across categories that not necessarily are driven by the intervention. This point should be</p>	<p>Drinks were classified as red, green and amber. Due to the reasons explained in Page 5, in the last paragraph of "Purchase data" subsection, we decided not to analyse the data from the amber drinks. Therefore, even if we chose to estimate the policy effect under a multivariate TS model, we would still not have covered the full profile of substitution. In</p>		

<p>incorporated in the analysis (i.e. estimate all time series jointly and allow free error correlation) or at least be explicit about the assumptions and limitations.</p>	<p>addition, we considered that univariate analysis would be easier to interpret. We have included this point as a limitation in the discussion (page 10).</p>
<p><b>6.</b> Given the way that the intervention was implemented, why not take an event-study approach? Using ITSA authors are obscuring the effects that are given by variation on the implementation strategies across centers (e.g. exclusion of diet carbonated drinks).</p>	<p>Thank you for this comment. The event-study approach is a common methodology predominantly used in the finance, econometrics, and strategic management literature. We choose to use interrupted time series analysis which is commonly used in evaluating public health interventions. We believe that variation on the implementation strategies across centres reflects the real-world conditions in which policies are applied and it has been taken into account fitting individual ITSA models (one per centre) and considering a model parametrization that allows for three different periods.</p>
<p><b>7.</b> Meta-analysis figures should all have the same order of centers in order to compare across different beverage categories. Moreover, is there any explanation of the substitution patterns observed within centers? this is key and largely ignored in the manuscript. Part of the success on implementation is not the reduction of "red" drinks but rather substitution towards healthier drinks, since consumer might easily substitute "red" drinks outside centers (e.g. at home). This should be discussed in detail.</p>	<p>Thank you for this comment, we have changed the order of the centres in Figures 4 and 5 to be consistent with Figure 3.</p> <p>In the original submission we briefly discussed substitution patterns and compared to existing literature: “..customers may have switched to the free, palatable drinking water available from centre water fountains, accounting for the lack of change in ‘green’ drinks” (page 11) and “Furthermore, Melbourne, Victoria has safe and palatable drinking water, and all centres had at least one accessible, free, water fountain. Settings and locations with less palatable and accessible water may see a more observable shift towards bottled water sales after SSB removal within their facilities.”(Page 11)</p> <p>We have now expanded on substitution effects within the limitations section of the discussion, page 10.</p> <p><i>A weakness of this study is the inability of sales data to represent actual beverage consumption and the lack of understanding of intervention effects on compensatory dietary behaviour in other settings. <u>Previous studies evaluating compensatory behaviour following removal of SSBs in school have found no increased consumption outside of school (30), or increased consumption that was of a smaller magnitude than the decrease in school-based consumption (31). While school settings are not completely analogous to sport and recreation settings, there is clear potential for availability changes in one settings to affect total consumption.</u></i></p>
<p><b>8.</b> Inter-rater reliability seems low. Authors should explain why.</p>	<p>Thank you for this comment. We revisited the inter-rate reliability analysis and examined the drinks that were miscategorised. The drinks that were miscategorised tended to be those that were not commonly available, and therefore unfamiliar to the coders, as well as having low numbers of</p>

	<p>sales. In total, the miscategorised drinks accounted for 0.49% of total sales over the four year period of analysis. Therefore, the low inter rater-reliability is due to drinks that are unlikely to change the outcomes of interest.</p> <p>We have included this information in Methods section page 5.</p>
<b>9.</b> For the reduction on drink varieties, baseline number should be provided as reference (or % change).	<p>Thank you for this suggestion, we have now included the average baseline drink varieties for red, amber, and green drinks in Results section, page 7.</p>
<b>10.</b> Following a previous point, why stratification by initiative implementation was not conducted? it seems critical to discuss results of the initiative.	<p>Thank you for this comment. As only one centre had fully implemented the policy, it was not possible to conduct a meaningful analysis stratifying by fully implemented or not. However, on reflection, we agree that this is an important analysis to conduct. Therefore, we assessed whether centres had removed the “core” SSBs, i.e. the carbonated sugar sweetened beverages. These beverages make up the bulk of the SSBs sold across centres.</p> <p>This information is now included in Table S5 in the Supplement.</p>
<b>11.</b> Why sales value matter? What do authors expect to infer from comparing change in volume versus monetary value? Similarly, when discussing change in the volume of sales (relative to total sales) there is a key problem with implicit correlation across groups being ignored in the estimation. Also, how were the CI calculated in this case?	<p>We use volume as the outcome when predicting the impact of the initiative on “healthy and unhealthy” sales of drinks, as this is a more direct measure of the potential health impact than monetary sales value. Dollar sales does not necessary reflect volume due to price promotions, different cost per litre depending on the size of the drink, etc.</p> <p>Sales value of drinks was used to assess the overall financial impact of the initiative. We acknowledge that this is an imprecise measure of financial impact but is likely to better reflect business outcomes than volume. Please note that we did not directly compare volume of sales relative to total sales.</p>
<b>12.</b> Saying that weighting does not affect the outcomes of meta-analysis is insufficient, specially when methods nor output is available on the supplement file.	<p>Thank you for this comment. We have described the methods in the Statistical Analysis section (page 5-6) and included Table S4 in the supplement reporting results of the analysis.</p>
<b>13.</b> Authors should be extremely careful with language. Sentences such as "ITSA is the strongest method for the evaluation of natural experiments" is extremely unfortunate and misguided.	<p>We have edited this sentence, page 10. “ITSA is a method for the evaluation of natural experiments [28,29] under a set of assumptions that...”</p>
<b>14.</b> How does ITSA mitigate the absence of a "control" center? it	<p>We agree with the reviewers that the absence of a control group is a real limitation to this analysis and that our</p>

<p>seems that authors are unclear about the assumptions underlying the method.</p>	<p>analysis relies on strong assumptions. As we now describe in the Statistical Analysis section (page 5-6) the SSB reduction initiative was imposed to all the centres within the YMCA at the same time and there was no opportunity to collect data in “control centres”. We have described the assumptions underlying the estimation of the initiative effect (Page 6) particularly the strong assumption for estimating the counterfactual outcomes.</p> <p>We have also edited the discussion to better reflect the limitations and strengths of ITSA approach, page 9.</p> <p>“A further limitation is the unavailability of “control” centres; as a consequence the estimation of the outcomes relies on the strong assumption that the pre-implementation model is valid into the future, i.e. there are no factors besides those included in the model that will affect the outcome and that the “effect” of these factors is constant across periods”</p>
<p><b>15.</b> Results suggest a decline of red drinks without further substitution. A natural approach should be a qualitative follow-up to understand whether this represent a total reduction on SSBs intake or substitution at home. As is, this study cannot argue to recommend this strategies as the claim above cannot be confirmed nor disprove (e.g. assuming that costumers are switching to water fountains is speculative). Overall, the conclusion is not supported by the results provided.</p>	<p>Thank you for this comment. We agree that a qualitative study would have complemented this work. However, this was not in the scope of this study.</p> <p>We have included previous research that has explored substitution effects in similar initiatives in the discussion, along with discussion regarding the use of drinking water fountains as possible substitution within the YMCA setting (p.9). We agree, that substitution effects of this intervention remain unknown and are an important focus of a future study.</p> <p>Previous research has suggested that children attending schools that remove SSBs do not substitute the full amount of SSB consumption from other settings. We have added this information into the discussion, page 9. Our recommendation of the strategy remains in terms of a strategy to reduce the sales of unhealthy drinks- the direct effect of the intervention.</p>
<p><b>Reviewer: 2</b> Reviewer Name: Penny Breeze</p> <p>Main limitations <b>1.</b> The statistical methods are not sufficiently described. This section should be re-written so that it is clear exactly what statistical models have been adopt</p>	<p>Following yours and other reviewer's recommendation we have re-written the Statistical Analysis section (page 5-6).</p>
<p><b>2.</b> The aims of the study and results are not consistent.</p>	<p>We have addressed this concern below (Results comment # 1)</p>

<p><b>Introduction</b>  <b>Intro. 1.</b> The introduction presents some background to the research in this area but excludes other very similar ITSA studies of sugar drinks reduction initiatives published. I was involved in similar study looking at price increases alongside promotion and product position changes in leisure centres in the UK also uses sales data and ITSA methods.</p>	<p>Thank you for this suggestion, we have included reference to this study on page 3. <i>“A further study using interrupted time series analysis to assess the impact of pricing changes to SSBs in UK recreation centres [21] offers methodological insights that could be applied to the evaluation of availability-based policies.”</i></p>
<p><b>Method</b>  <b>Meth 1.</b> More information is needed to describe the type of statistical model used. It is not sufficient to refer to the stata command used. Ideally the methods should be described in more detail within the main manuscript, with additional detail provided in the appendix. More detail is needed on the methods used for the meta-analysis and a justification given for equal weighting.</p>	<p>Thank you for this comment. We have expanded the Statistical section to include more details on pages 5 and 6.</p>
<p><b>Meth 2.</b> More detail needs to be provided on the choice of methods for this analysis. Meta-analysis is not the only method available for allowing heterogeneity between study site so it needs justification.</p>	<p>Thank you for this comment. The centres included in this study vary in a number of key ways that are likely to affect the purchase of SSBs and other drinks, and result in large outcome variability in both size and time series profile between centres. We have included Figure S2 in the Supplement which show that volume of drinks sold varies across centres. Furthermore, centres have different facilities (e.g. only some have swimming pools), which mean that attendance and seasonality may have different impacts on the outcomes of interest. Because of these characteristics of the setting under study, and therefore outcome measures, we ruled out an approach such as linear mixed model, or panel mixed method, which rely on assumptions such as homogeneity of season and attendance impact across centres. As we now explain in the Statistical Analysis section (Analytical approach and rationale) we opted for fitting the segmented regression model for each centre individually and estimating the relative change in the outcome of interest, and combined the results in a meta-analysis.</p>
<p><b>Meth 3.</b> I agree with adjustment for seasonal changes in sales for the analysis. However, grouping months by season may not be the most appropriate way of defining these changes. For example one might</p>	<p>Thank you for this suggestion. We have incorporated examples of the raw data in Figure S2 in the Supplement to illustrate the profile of sales in 2 centres. Seasonality and attendance are adjusting for slightly different concepts. Seasonality adjusts for general consumption patterns, e.g. higher consumption per person in hotter months, while</p>

<p>expect that months with school holidays may see peaks in attendance independent of weather related changes. It would be useful to provide some descriptive statistics of sales data to justify your approach.</p>	<p>attendance adjusts for the higher probability of drinks being purchased the more people that are in the centre, e.g. there are more people in the centre because of an unusual hot week in winter, a sport event or school holidays. We hypothesize that both factors contribute to purchases of drinks.</p> <p>We have included a summary of volume of drinks sales across the whole study period and plots showing the trend in two centres. Unfortunately due to commercially sensitive of the data we can't provide more detailed information on sales at the centre level or all the plots.</p>
<p><b>Meth 4.</b> It would be useful to see some illustration of the magnitude of sales, fluctuations over time and how they change during the pre, implementation and post-implementation period.</p>	<p>Thank you for this suggestion, we have included raw data on 'red' volume sales in Figure S1 for two centres as well as identifying the pre-implementation, implementation, and post implementation periods. The two centres are include one centre with a swimming pool and moderate drink sales, and a centre with a stadium and high drink sales.</p>
<p><b>Meth 5.</b> It is not clear to me how the timing of the initiative work. Firstly, did all centres start the initiative at the same time or does the post-initiative period vary by centre? Secondly, it is not clear why the period Dec 2015-Dec 2016 was excluded from the primary analysis. This is particularly important in looking at how people adapt to the changes in drinks provision over time. I can see the benefit of looking at the two endpoints separately, but would prefer to see both within the main text.</p>	<p>We have modified the text to better describe the initiative (page 4) and updated Figure 1. In the new version of Statistical Analysis we now present the model that we fitted for the different analyses, which include the full time series data to estimate the relative change in the outcomes of interest at both time points.</p> <p>We have chosen these time periods, as the centres were given from the start of December 2014 until the start of December 2015 to implement the SSB reduction initiative. We chose to describe three time periods as we were aware that the rate of implementation was likely to differ across the centres. Further, it was likely that some centres had not in fact been able to full implement, or maintain, the SSB reduction initiative. We chose the third time period (December 2015-December 2016) as December 2015 was the date by which the initiative was intended to be implemented. Assessing the impact of the SSB reduction initiative on an intention to treat basis was appropriate given the importance of assessing the initiatives effects in a "real-world" setting. We further chose to assess the initiative outcomes at December 2016 to gain insight into the sustained impact, one year following expected implementation. We have expanded on this in the text, pages 5 and 6.</p> <p>Unfortunately we are only able to present 5 figures in the main text. We have included an additional Table in the Supplement S4 which allows for direct comparison between the two timepoints.</p>



<p><b>Meth 6.</b> It is not conventional to reference results from the supplementary material in the methods section. If sensitivity analyses are performed they should be described in full within the methods or supplementary material.</p>	<p>Thank you for this comment. We have moved supplement material into the methods section pages 5 and 6.</p>
<p><b>Results</b> <b>Results 1.</b> Given the objective of the study I am surprised to see the first section of the results address whether the initiative was successfully adopted by the centres. It may be necessary to rethink what is reported in the results section or re-word the aims of the study.</p>	<p>Thank you for this feedback. We aimed to assess the impact of this real world implementation on sales outcomes. We have clarified in the text that we conducted an intention to treat analysis – that is, assessing the outcomes of an organization initiative that was the directive of the head office of YMCA, whether or not the initiative had been fully implemented in each centre. We believe it is important to present the degree of implementation in order to fully represent what has occurred. We have clarified our approach in the text on page 5.</p>
<p><b>Results 2.</b> It is not clear how the availability of drinks was assessed. I was under the impression that the study had access to sales data. Is it possible that there is wastage at the sites if products are not sold? Wouldn't this affect the accuracy of the availability analysis?</p>	<p>Availability was assessed using sales data. We have updated the methods to clarify this, page 5-6. Sales data is an imperfect measure of availability, as it is possible that there were some products that were available but didn't sell during the month. However, this was the best available measure that we could access. We have included this information in the discussion of limitations of the study (Page 10).</p>
<p><b>Results 3.</b> The statement of inter-rater reliability is not suitably placed and requires further comment to make it clear what this refers to. Does this mean that there was some variation between centres as to how products were categories into the red-amber-green groups? If so the level of disagreement is surprising. Can you identify which types of products were most likely to be mis-categorised? From the methods the products were categories by a dietitian, was this done by the research team or locally by the centres to implement the policy?</p>	<p>Thank you for this comment. Please see response to Reviewer 1- comment 8.</p>
<p><b>Results 4.</b> Given that the centres did not introduce the policy by Dec2015 I am even more surprised that the data you have beyond this</p>	<p>As stated above, for all analyses we fitted a model that include the full time series data (Jan 2013 – March 2017). The new version of the Statistical analysis section includes the model.</p>

period has not been used in the primary analysis.	
<b>Results 5.</b> I think the results section would benefit from a table or figure illustrating the variation in results between the Dec 2015 and Dec 2016 end-points.	Thank you for this recommendation, we have included a Table that demonstrates the outcomes at both timepoints, see Table S4.
<b>Results 6.</b> “Stratification by variables of interest” is ambiguous. What are the variables of interest?	We have clarified this in the text, page 6.
<b>Discussion</b> <b>Disc 1.</b> I think the discussion of how the results compare with other studies could be improved. The manuscript refers to an Australian and Canadian study but provides no information on whether the policies implemented were similar to this one.	Thank you for this comment, we have provided further detail in the text, page 10-11.
<b>Disc 2.</b> It is interesting to note that a similar study in the UK found that a SSB initiative decreased sales of SSB, and increased sales of diet carbonated drinks, but did not observe changes in healthy options (fruit drinks or water).	Thank you for this suggestion. We have included this important reference that was omitted from the original submission that has improved the discussion, page 11. <i>“Our findings of a lack of increase in “green” sales is similar to a study examining the impact of a 20% price increase on SSBs in seven UK recreation centres where patrons switched from the unhealthier SSBs to artificially sweetened SSBs, rather than to water [21]. “</i>
<b>Reviewer: 3</b> Reviewer Name: Cordia Chu It would be helpful for the authors to explain clearly the intervention strategies applied by the 16 Australian recreation centres for the sugar-sweetened beverage reduction initiative.	We have clarified the description of the initiative, see page 4.
<b>Reviewer: 4</b> Reviewer Name: Beatrice Biondi  As a general comment, the paper looks promising; data and study settings can form the base for a valuable and sound research. However, the application of the methodology and the displaying of results is not straightforward. From my point of view – which is the one of a statistician working with	We address each of the comments below.

<p>quasi-experimental methods, but not in the field of public health nor of clinical studies – the paper lacks in explanations and justifications for the researchers' choices of the statistical analysis, which affect the reliability of final results.</p> <p>I provide below some comments, with specific reference to the implementation and presentation of statistical methods in the paper, which I believe can improve the research overall. Points 1-5 represent the main weaknesses of the study, the other points are minor revisions in text.</p>	
<p><b>1. Definition of the intervention:</b> after an in-depth reading, I understood that the original aim of what you called a reduction initiative was instead a complete removal of "red" SSB, and recreation centres applied it to a different degree (only one removed all red SSB). At page 4, lines 15-20, you explain the aim of the initiative, providing a very clear table for classification of beverages. Here, you should also clarify that not all centres applied this to the same extent. This is extremely important also for the assumptions behind the model: this means that the intervention is somehow different for all centres. You must provide descriptive statistics on the share of red SSB reduction implemented by each centre. And it is essential to know whether and when the change in assortment occurs: in order to estimate the effect of the reduction, assortment should (I) remain nearly stable in the pre-intervention period, (II) change with the intervention, (III) remain again nearly stable in the post-intervention period.</p>	<p>Thank you for this comment. We have clarified the description of the initiative (page 4), which is the removal of all "red" SSBs excluding sports drinks.</p> <p>We have also added information that not all centres implemented the initiative on page 4 where we add further detail on the implementation measures presented:</p> <p>"We have also included Table S3 which presents centre-level results of change in 'red' and 'green' volume and share of red varieties at two timepoints. Unfortunately the data available to us does not allow to identify when the change in assortment occurred.</p>
<p><b>2. Gradual implementation:</b> the removal of red SSB was implemented gradually during the implementation year, at least this</p>	<p>Thank you for your comment. We believe that our approach provides valid estimates for the policy effect at centre level (when all the underlying assumptions are valid). Our model includes three periods to acknowledge that the intervention</p>

<p>is what I understood from the text. This could create problems in the estimation of ITS. Indeed, ITS “is not appropriate when (...) the intervention is introduced gradually or at more than one time point” (Kontopantelis <i>et al.</i>, 2015), you should control for this issue and provide evidence of how reduction share changes at the beginning and at the end of the implementation period.</p>	<p>was gradually introduced, and we fitted individual models for each centre to avoid the strong assumption that all the centres followed the same temporal trend (conditional on season and attendance). Our model strategy is accounting for the fact that different centres had different level of initiative uptake. Further we have now clarified that the average initiative effect across centres should be understood as an ITT estimate of the initiative effect, i.e. the effect of the head office directing centre managers to implement the initiative.</p>
<p><b>3. Period of implementation:</b> How you considered pre- and post- intervention periods is not at all clear in the paper. In the ITS procedure you should identify a cut-off time point to divide the observations in pre and post, and it is confusing to me whether you take Dec 2014 or Dec 2015. At page 4, lines 5-13 you say that the intervention is one action of a series of actions, belonging to a broader policy, implemented over a three years period. Which period? Then you say that the intervention has been implemented over one-year period, completed by December 2015. Fig 1 then is divided in three periods, and the reader is unsure about the pre/post intervention time span. Finally, going to the methods section, you clearly say you used three periods. But what does change between implementation (Dec 2015) and post-implementation (Dec 2016)? Did the gradual removal that happened along 2015 reach a final point (share of red SSB) and remain constant in 2016? I would carefully reconsider the choice of choosing three periods, since it means two shocks took place: for instance, in their example Linden and Arbor (2015) consider the cases in which an intervention is introduced, withdrawn, and reintroduced; or an intervention is followed by a separate intervention at a later point in time. It seems to me that in your case only one intervention was</p>	<p>We have further clarified the intervention, page 4.</p> <p>We preferred to allow for two additional parameters in the model to account for the fact that by the beginning of December 2015 all centres should have the initiative implemented. Even if there wasn't a shock, it actually indicates a deadline that was given to the centres to have the initiative in place. In theory, for 'red' drinks we expected a negative trend during the implementation period and a different trend, probably no trend, in the post-implementation period. We considered that a less restrictive model, with two more parameters, would be less likely to be mis-specified. We have clarified our approach and justification in the text, page 5-6.</p>

<p>considered. This could also explain some anomalous results you obtain (see point 14).</p>	
<p><b>4. Consideration of “amber” products:</b> on page 5, lines 3-7 is said that amber products are not considered because some centres removed them and others reduced it. I think this is not a reason to not consider this type of products. In centres that implement a reduction, you could consider this type of products in the same way of red products. At page 9, line 53 you say the analysis is precluded. Instead, you could run an analysis on the 7 centres that sold amber products, and analyze purchase patterns focusing on substitution.</p>	<p>Thank you for this suggestion. Unfortunately, the sales volume of amber in each of the centres that still had amber products was very low on a monthly basis (tended to represent approximately 7-10% of sales). These low numbers produced high month to month variability that prevented meaningful results to be produced. We have now added this explanation in the text on page 5.</p>
<p><b>5. ITSA specification:</b> Linden and Arbor (2015): “when the treated group’s outcomes can also be contrasted with those of one or more comparison groups, the internal validity is further enhanced by allowing the researcher to potentially control for confounding omitted variables.” and Linden (2018) “When ITSA is implemented without a comparison group, the internal validity may be quite poor. Therefore, adding a comparable control group to serve as the counterfactual is always preferred.” Given these, I was surprised to see that you did not use a control group. Especially given that only 8 out of 68 centres have not accessible sales data (I assume the other 60 have accessible data), and you explicit that 5 centres did not commit to remove SSB. They could potentially be a perfect control group. At page 9, line 27 you say that control centres are unavailable, why is that? I encourage you in performing the analysis using the control group. You could use simple multiple group ITSA (Linden, 2015), or use a comparable control group,</p>	<p>Thank you for your comment. We agree that the presence of a control group would enhance the validity of the findings. As we now describe in the Statistical Analysis section (page 5-6) the SSB reduction initiative was imposed to all the centres within the YMCA at the same time and there was no opportunity to collect data in “control centres”.</p> <p>We have described the assumptions underlying the estimation of the initiative effect (Page 5-6) and have also edited the discussion to better reflect the limitations and strengths of ITSA approach, page 9.</p>

e.g. by matching on covariates related to centres (see Linden, 2018). You could consider also other methods, such as difference-in- difference. In any case, I believe the study results would greatly gather in reliability if you use a control group in the estimation.	
<b>6.</b> You did not explain how your “novel meta-analysis” is performed. I assume you averaged across centres. Did you explore the possibility of running a panel fixed-effect model, instead of running 16 different models? This might represent a more “elegant” solution, which allows to consider the difference among centres.	Thank you for this comment. Please refer to the response to Reviewer 2 Methods 2 query.
<b>7.</b> Price changes are not controlled for in the model. Did you check for changes in prices in the period considered? This could greatly affect your results.	Thank you for this recommendation. We revisited the original data to examine whether there were price changes. We observed small and consistent increases in price per unit over time across all centres (less than 4% annually). We have added this detail to the manuscript
<b>8.</b> Results section: A graph of the trend in volume sales for red SSB on the period considered, cumulative for all the centres or even with a line for each centre, would help the reader in understanding the overall change in consumption over time.	We have now included graphs from different types of centres to demonstrate overall changes in consumption over time, see Figure S2 in the Supplement.
<b>9.</b> Page 7, line 38 here you say S6 table but you refer to S7; line 42 if I look to S5 table I see a 57.2 increase in green sales. How do you explain this increase after a 0 increase in 2015? At line 43 again you should refer to S7 table, where I see some differences among strata, indeed. Reformulate this paragraph according to the results shown in tables.	Thank you for this comment. We have corrected the Table names and references. The 57.2% increase is not significant (CI -46.1, 160.6), and therefore consistent with the 12 month results.
<b>10.</b> Discussion Section: Page 9 lines 18-21: why do you think is best to avoid having all centres in one model? Please, motivate this sentence.	Thank you for this comment. We refer to the reviewer’s comment 5, where we respond with arguments for the use of ITSA, which is based on the variability in sales data across centres.



<p><b>References:</b>Kontopantelis, E. <i>et al.</i> (2015) 'Regression based quasi-experimental approach when randomisation is not an option: Interrupted time series analysis', <i>BMJ (Online)</i>, 350, pp. 1–4. doi: 10.1136/bmj.h2750.</p> <p>Linden, A. (2018) 'A matching framework to improve causal inference in interrupted time-series analysis', <i>Journal of Evaluation in Clinical Practice</i>, 24(2), pp. 408–415. doi: 10.1111/jep.12874.</p> <p>Linden, A. and Arbor, A. (2015) 'Conducting interrupted time-series analysis for single- and multiple-group comparisons', <i>Stata Journal</i>, 15(2), pp. 480–500. doi: 10.1177/1536867X1501500208</p>	
---	--

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Juan Caro Health Policy and Management Department University of North Carolina at Chapel Hill North Carolina, USA
<b>REVIEW RETURNED</b>	06-Oct-2019

<b>GENERAL COMMENTS</b>	<p>I appreciate the opportunity to revise a new version of this research article. The manuscript has improved significantly. However, I still have very critical concerns.</p> <p>First, from a public health and policy perspective, there is no justification in the study (or elsewhere) to consider total sales a relevant outcome, unless a cost-benefit analysis is reported (consumer welfare and producer surplus).</p> <p>Similarly, the discussion indicates that the study demonstrates feasibility, acceptability and effectiveness, while the analysis is only impact on purchases within centers. There is a large literature on beverage substitutions across points-of-sale, which is ignored in this paper. Moreover, back-of-envelope calculations regarding SSB reductions per individual are misleading; in addition to substitutions with other unhealthy beverages elsewhere and the fact that clients do not attend centers everyday, the change in purchases can be driven only by a small fraction of high-consumers, as noted in the literature.</p> <p>In general, the discussion over-reaches and it is not balanced with the presented evidence. In fact, no changes in healthy beverages only support the hypothesis of unhealthy beverage substitution in other settings.</p>
-------------------------	--

	Finally, regarding the supplement, the presentation of the data is misleading; if not presenting the full trend of all centers (weighted and unweighted), trends should be presented for all centers. In fact, the presented evidence (figure 1) is incomplete (e.g. what does "fitted" mean), and moreover it shows no change in one of the centers as is. I suggest to remove the seasonal component before presenting the figures in order to clearly show that the ITSA assumptions hold.
--	---

<b>REVIEWER</b>	Penny Breeze University of Sheffield, United Kingdom
<b>REVIEW RETURNED</b>	09-Oct-2019

<b>GENERAL COMMENTS</b>	<p>Many thanks for your response to the reviewer comments. My main concerns with the paper have been addressed. The study provides a really interest contribution to the literature on approaches to reducing sugar intake. I have a few further comments on the manuscript.</p> <ol style="list-style-type: none"> <li>1. I found the stratified analysis by implementation hard to follow within the manuscript. I think this could be defined more clearly in the methods section. It would also be useful to interpret these results with reference to the proportion of 'red' sales that were carbonated drinks at baseline.</li> <li>2. I found the reporting of changes in sugar purchased insufficient. I realise that you are limited by what can be included in the manuscript but I do not see why the meta-analysis cannot be included in the supplementary material. It is not at all clear how the calculation of 577 kilograms of sugar is derived. This result is particularly difficult to interpret because the paper does not report total sugar consumption.</li> <li>3. The exclusion of 'amber' products would benefit from further justification. Couldn't these have been incorporated into green products because they are low sugar. Alternatively it seems like the leisure centres tended to classify them as unhealthy and reduce their availability, in which case they could have been included as 'red'. Diet carbonated drinks are an consideration when looking at substitution effects, so the exclusion of this data seems incomplete. It would be useful to know if they impact on the findings in a sensitivity analysis.</li> </ol>
-------------------------	--

<b>REVIEWER</b>	Beatrice Biondi University of Bologna, Italy
<b>REVIEW RETURNED</b>	15-Oct-2019

<b>GENERAL COMMENTS</b>	<p>It seems that the authors did a lot of work in reframing the text to comply with reviewers' concerns. However, minor changes to analysis and methods have been performed, and sometimes the answers to the comments do not really solve them.</p> <p>In general, I still believe that the analysis method is not adequate and/or the research question is not correctly stated. The authors should decide whether defining the initiative as a "reduction" or a "removal", currently, they use both the terms and this is unclear. If the aim of the initiative was to remove completely the targeted red drinks, it makes no sense to analyze the trend in the red drinks purchases in the post-implementation period (it should be zero). If the centres adopted the removal to a different degree, it should be</p>
-------------------------	---

	<p>presented in the data (see my previous comment 1), not as a result, but as a description of your data.</p> <p>Another major flaw concerns the approach used. The authors say that “All analysis were conducted under an intention to treat approach, i.e. we assessed the impact of the initiative regardless of the extent of initiative implementation.”. This seems a misleading interpretation of the ITT approach. Looking at the literature, “Intention to treat is a strategy for the analysis of randomized controlled trials that compares patients in the groups to which they were originally randomly assigned.” (Hollis &amp; Campbell, 1999). No randomization was made in this study, therefore referring to this type of analysis is inaccurate.</p> <p>Comment 5 about control group: the authors say that “Given that the SSB reduction initiative was introduced from head office at the same time in the entire organization there was no opportunity to collect data in “control centres”. But looking at Table S3, centres 6,7 and 12 did not decrease number of red drinks available, and centres 4 and 10 only removed one product. Couldn’t these be used as control centres? Why are they included in the analysis if they did not decrease availability of red drinks? This is unclear.</p> <p>Comment 8: I do not see Figure S2 in the Supplement. If the authors refer to Figure S1, this is not what I was asking for, since only two groups are presented. More importantly, I cannot figure out why there is still a trend in post-implementation period, since all the red drinks should have been removed. No sales should be observed in this period. Otherwise the initiative has not been applied correctly and this has to be specified.</p> <p>Comment 9: How did you calculate CI? It is quite striking to me that a 57% increase could be not significant... Indeed your CI are sometimes too large to be interpreted.</p> <p><b>METHODS</b></p> <ul style="list-style-type: none"> <li>• “which individual centres were given one year, from December 2014 to November 2015, to implement”, a more readable form would be “which was gradually implemented by centres over a period of one year, from December 2014 to November 2015”.</li> <li>• “Figure 1 demonstrates the periods involved in the SSB reduction initiative”. I would say the figure does not “demonstrates the period”. Why did you choose only two centres and how did you select them? The legend cannot be understood from the reader: how are the centres fitted? How is the counterfactual estimated? There is no explanation regarding what Figure 1 shows throughout the paper. Some explanation is provided to reviewer 1- comment 4, this needs to be improved and included in the manuscript since the reader will likely not read the review document</li> </ul> <p>I believe the analysis should be re-thought to take into account the peculiar structure and characteristics of the dataset.</p>
--	--

## VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:	Author responses
<p><b>Reviewer: 1</b></p> <p><b>Reviewer Name: Juan Caro</b></p>	

<p>Institution and Country:</p> <p>Health Policy and Management Department</p> <p>University of North Carolina at Chapel Hill</p> <p>North Carolina, USA</p> <p>Please state any competing interests or state 'None declared': None declared</p> <p>Please leave your comments for the authors below</p> <p>I appreciate the opportunity to revise a new version of this research article. The manuscript has improved significantly. However, I still have very critical concerns.</p>	<p>We have addressed each concern individually below.</p>
<p>First, from a public health and policy perspective, there is no justification in the study (or elsewhere) to consider total sales a relevant outcome, unless a cost-benefit analysis is reported (consumer welfare and producer surplus).</p>	<p>Thank you for this comment. We consider "total sales" as proxy for "profitability". Total sales has been used previously in this way in at least 25 studies as identified in a systematic review (Blake et al, 2019).</p> <p>We have added information and reference in the text to this point, page 6:</p> <p><i>"Dollar sales of drinks is used as a proxy for profitability and to assess the potential financial impact of the initiative."</i></p> <p>We have further added the following sentence to the limitations: (page 11)</p> <p><i>"Total sales is an imperfect proxy for profitability and financial viability, however the reporting of business outcomes is important when considering the application of policies to commercial settings."</i></p> <p>Our work with food retailers (Boelsen-Robinson et al, 2018; Blake et al., 2017; Boelsen-Robinson et al, 2017) and our recent review (Blake et al, 2019) demonstrate that feasibility and sustainability of healthy food retail interventions is often dependent on the perceived viability by the retailer. Total sales is one indicator used by retailers to assess this viability (Boelsen-Robinson et al, 2018; Blake et al., 2017; Boelsen-Robinson et al, 2017; Blake et al, 2019). We would argue that given the emerging nature of this research field, building data that is relevant to the intervention implementers as well as the public health community is important in order to drive the field forward.</p> <p><u>References</u></p> <p>Blake MR, Peeters A, Lancsar E, Boelsen-Robinson T, Corben K, Stevenson CE, et al. Retailer-Led Sugar-Sweetened Beverage Price</p>

	<p>Increase Reduces Purchases in a Hospital Convenience Store in Melbourne, Australia: A Mixed Methods Evaluation. Journal of the Academy of Nutrition and Dietetics. 2017.</p> <p>Blake MR, Backholer K, Lancsar E, Boelsen-Robinson T, Mah C, Brimblecombe J, et al. Investigating business outcomes of healthy community food retail strategies: a systematic scoping review. 2019.</p> <p>Boelsen-Robinson T, Backholer K, Corben K, Blake MR, Palermo C, Peeters A. The effect of a change to healthy vending in a major Australian health service on sales of healthy and unhealthy food and beverages. Appetite. 2017;114:73-81.</p> <p>Boelsen-Robinson T, Blake MR, Backholer K, Hettiarachchi J, Palermo C, Peeters A. Implementing healthy food policies in health services: A qualitative study. Nutr Diet. 2018.</p>
<p>Similarly, the discussion indicates that the study demonstrates feasibility, acceptability and effectiveness, while the analysis is only impact on purchases within centers. There is a large literature on beverage substitutions across points-of-sale, which is ignored in this paper. Moreover, back-of-envelope calculations regarding SSB reductions per individual are misleading; in addition to substitutions with other unhealthy beverages elsewhere and the fact that clients do not attend centers everyday, the change in purchases can be driven only by a small fraction of high-consumers, as noted in the literature.</p> <p>In general, the discussion over-reaches and it is not balanced with the presented evidence. In fact, no changes in healthy beverages only support the hypothesis of unhealthy beverage substitution in other settings.</p>	<p>Thank you for this comment. We agree acceptability was not measured and we have removed this point.</p> <p>We have amended this paragraph to make our intent clearer (page 13):</p> <p><i>“Studies such as this demonstrating that healthy food and beverage interventions can be feasible and effective to implement, whilst only moderately impacting overall sales within these organisations can aid a shift to healthy food retail more broadly by lowering the perceived risk.”</i></p> <p>Here we are referring to effectiveness of the implemented intervention, which resulted in a decrease in unhealthy drinks purchased. We agree this study does not address changes in overall diet, but measures direct effectiveness of a retail intervention on healthiness of purchasing from that retail outlet.</p> <p>We have made this clearer, amending the following paragraph in the discussion (Page 12).</p> <p><i>“Replacing just one can of SSB a day with water can significantly improve health and has been modelled to result in a modest but significant reduction in obesity rates. Extrapolation of the response to this policy to all sports and recreation facilities across Australia would lead to over 3.5 million fewer cans of SSB purchased per year. Future research is needed to analyse the impact of the purchasing reductions of healthy food retail interventions such as this on overall diet.”</i></p> <p>We have also added this paragraph to the discussion to acknowledge the potential substitution to SSBs outside of sport centres (page 12):</p>

	<i>Customers may seek to purchase unhealthy drinks elsewhere as a result of low availability within sports centres. The evidence on the degree to which they are likely to do is limited and mixed (34,35, 40). Future research is needed to identify the extent of such substitution and strategies to reduce it. Government policy to reduce SSB availability across more settings is one means of addressing compensatory behaviours, and creating a level playing field where retailers face fewer disadvantages for taking action on public health nutrition.</i>
Finally, regarding the supplement, the presentation of the data is misleading; if not presenting the full trend of all centers (weighted and unweighted), trends should be presented for all centers. In fact, the presented evidence (figure 1) is incomplete (e.g. what does "fitted" mean), and moreover it shows no change in one of the centers as is. I suggest to remove the seasonal component before presenting the figures in order to clearly show that the ITSA assumptions hold.	<p>Figure S1 showed the raw and the estimated time series (counterfactual and under intervention) for two of the centres. We have clarified in the footnote of the figure that the "raw" time series are the original data with no adjustment, so that the reader could appreciate that the model fitted the data reasonably well.</p> <p>The centres in Figure S1 were selected as examples of how the volumes sales can vary across different types of centres (centres with different facilities).</p> <p>We have now included an additional set of Figures in the Supplement showing the raw data for "red" drinks sales, i.e. unadjusted, with the centres grouped by volume sales, see Figure S2.</p>
<p><b>Reviewer: 2</b></p> <p><b>Reviewer Name: Penny Breeze</b></p> <p>Institution and Country: University of Sheffield, United Kingdom</p> <p>Please state any competing interests or state 'None declared': Non declared</p> <p>Please leave your comments for the authors below.</p> <p>Many thanks for your response to the reviewer comments. My main concerns with the paper have been addressed. The study provides a really interest contribution to the literature on approaches to reducing sugar intake. I have a few further comments on the manuscript.</p>	<p>We have addressed the comments individually below.</p>
<p>1. I found the stratified analysis by implementation hard to follow within the manuscript. I think this could be defined more clearly in the methods section.</p> <p>It would also be useful to interpret these results with reference to the proportion of 'red' sales that were carbonated drinks at baseline.</p>	<p>Thank you, we have clarified the paragraph in the text in the Statistical Analysis section page 7.</p> <p>Thank you for this suggestion. However, we believe it may be confusing for the reader if we were to compare descriptive proportions of volume purchases with percentage reduction in volume derived from the ITSA.</p>



2. I found the reporting of changes in sugar purchased insufficient. I realise that you are limited by what can be included in the manuscript but I do not see why the meta-analysis cannot be included in the supplementary material.	The estimated impact of the intervention on sugar sold in packaged drinks was reported in Table S4 in the supplementary and reference to these results was included in the text. As pointed out by the reviewer not all information can be included in the main paper. Please let us know if you are referring to further information.
It is not at all clear how the calculation of 577 kilograms of sugar is derived. This result is particularly difficult to interpret because the paper does not report total sugar consumption.	We have added in page 7 further explanation in the methods section for the sugar purchase estimation.
3. The exclusion of 'amber' products would benefit from further justification. Couldn't these have been incorporated into green products because they are low sugar. Alternatively it seems like the leisure centres tended to classify them as unhealthy and reduce their availability, in which case they could have been included as 'red'. Diet carbonated drinks are an consideration when looking at substitution effects, so the exclusion of this data seems incomplete. It would be useful to know if they impact on the findings in a sensitivity analysis.	As the reviewer indicates, it is not clear whether the amber drinks would be more appropriately categorised together with green "healthier" or red "unhealthy". We also agree that diet carbonated drinks are a consideration when examining substitution effects – however we feel that the unusual context where diet drinks were removed in some centres and not others means we would be misrepresenting the change in either 'green' or 'red' drinks if amber drinks were to be combined with either of them.  We have clarified in the text (page 5) that only 6 centres had sold diet soft drinks for the duration of this study. We believe that a sensitivity analysis on only 6 centres would not provide any additional insights into purchasing behaviours as a result of this initiative.
<b>Reviewer: 4</b> <b>Reviewer Name: Beatrice Biondi</b> Institution and Country: University of Bologna, Italy Please state any competing interests or state 'None declared': None declared Please leave your comments for the authors below It seems that the authors did a lot of work in reframing the text to comply with reviewers' concerns. However, minor changes to analysis and methods have been performed, and sometimes the answers to the comments do not really solve them.	We have addressed reviewer comments individually below.
In general, I still believe that the analysis method is not adequate and/or the research question is not correctly stated. The authors should decide whether defining the initiative as a "reduction" or a "removal", currently, they use both the terms and this is unclear. If the aim of	Thank you for this comment. We have attempted to address this comment previously, by referring to the initiative as "SSB reduction initiative" throughout the manuscript. However, it is accurate that certain product categories (e.g. soft drink) were expected to be "removed" rather than

<p>the initiative was to remove completely the targeted red drinks, it makes no sense to analyze</p> <p>the trend in the red drinks purchases in the post-implementation period (it should be zero). If the centres adopted the removal to a different degree, it should be presented in the data (see my previous comment 1), not as a result, but as a description of your data.</p>	<p>“reduced”, and we would like to retain this wording for accuracy. We understand that the initiative is complex in nature and have further edited the description of the initiative to clarify (page 4).</p> <p><i>“The initiative aimed to reduce the availability of packaged SSBs classified as ‘red’. This was done by focusing on the complete removal of some types of ‘red’ SSBs, and the reduction of the availability of other types of ‘red’ SSBs. Across the centres there was variability in the extent to which the intended removal and reduction of ‘red’ drinks was achieved. Products intended for complete removal included non-diet carbonated beverages, flavoured water, high kilojoule flavoured milk, and fruit drinks with less than 99% fruit juice. Fruit juice &gt;250ml are also classified as ‘red’ and were intended for removal. Sports drinks were to be reduced to &lt;10% and ‘green’ options increased to &gt;70% of fridge space. Other ‘red’ drinks such as SSBs that were prepared onsite were not focus of the initiative (e.g. milkshakes, see S1 Table in the supplement) and were still available for purchase.”</i></p>
<p>Another major flaw concerns the approach used. The authors say that “All analysis were conducted under an intention to treat approach, i.e. we assessed the impact of the initiative regardless of the extent of initiative implementation.”. This seems a misleading interpretation of the ITT approach. Looking at the literature, “Intention to treat is a strategy for the analysis of randomized controlled trials that compares patients in the groups to which they were originally randomly assigned.” (Hollis &amp; Campbell, 1999). No randomization was made in this study, therefore referring to this type of analysis is inaccurate.</p>	<p>Thank you for this comment.</p> <p>We agree with the reviewer that the classical RCT definition of “intention to treat analysis” does not apply in this context. However we chose to use this terminology because it clearly implies that we aimed to measure the effect of the initiative indication (similar to treatment indication in an RCT) regardless of whether or not the centres adhered to the initiative or there were initiative deviations or compliance concerns.</p>
<p>Comment 5 about control group: the authors say that “Given that the SSB reduction initiative was introduced from head office at the same time in the entire organization there was no opportunity to collect data in “control centres”. But looking at Table S3, centres 6,7 and 12 did not decrease number of red drinks available, and centres 4 and 10 only removed one product.</p> <p>Couldn’t these be used as control centres? Why are they included in the analysis if they did not decrease availability of red drinks? This is unclear.</p>	<p>Thank you for this comment.</p> <p>Because all the centres were part of the YMCA organization who proposed the initiative, using as “controls” those centres that decided not to adhere to the policy would imply that we are comparing “compliers” to “non-compliers” which are very likely to differ in characteristics that have not been measured in our study.</p> <p>Again, our goal was to describe the effect of an initiative indication, which was not strictly enforced, in a recreational sport setting. We have acknowledged the limitations of not having control centres in Discussion section (page 11).</p>
<p>Comment 8: I do not see Figure S2 in the Supplement. If the authors refer to Figure S1,</p>	<p>Our apologies, we were indeed referring to Figure S1. We have now included raw data for all centres in the supplement, Figure S2.</p>

<p>this is not what I was asking for, since only two groups are presented.</p>	
<p>More importantly, I cannot figure out why there is still a trend in post-implementation period, since all the red drinks should have been removed. No sales should be observed in this period. Otherwise the initiative has not been applied correctly and this has to be specified.</p>	<p>Thank you for this comment. As indicated in the first section of the results, most centres did not fully implement the initiative by December 2015 (i.e. reduce SSBs to only sports drinks), see page 9 and below. This meant that centres continued to remove SSBs after December 2015 (expected initiative completion date).</p> <p>Below is the paragraph where this is stated, page 9.</p> <p><u>Initiative Implementation</u></p> <p><i>“Fourteen of the sixteen centres had a reduced number of ‘red’ cold packaged drinks for sale by December 2015, however only one centre had fully implemented the initiative (i.e. did not sell a non-sports drink ‘red’ cold packaged drink at this time point).”</i></p>
<p>Comment 9: How did you calculate CI? It is quite striking to me that a 57% increase could be not significant... Indeed your CI are sometimes too large to be interpreted.</p>	<p>Thank you for this comment.</p> <p>As explained in the Methods section (page 6-7), the percentage change <math>\Delta_t</math> for each centre was estimated as a non-linear combination of parameters <math>\Delta_t = \frac{(\mu_{int,t} - \mu_{cf,t})}{\mu_{cf,t}}</math>. The parameters <math>\mu_{int,t}</math> and <math>\mu_{cf,t}</math> were predictions obtained under the models described in Eq1 and Eq2 respectively. The CI for <math>\Delta_t</math> was estimated using the Delta Method.</p> <p>Wide confidence intervals were associated with poor model fitting (large residuals) for some centres and the fact that the counterfactual prediction is well beyond the range of data used to fit the model. The last issue is particularly relevant for confidence intervals calculated for December 2016.</p>
<p>METHODS</p> <ul style="list-style-type: none"> <li>• “which individual centres were given one year, from December 2014 to November 2015, to implement”, a more readable form would be “which was gradually implemented by centres over a period of one year, from December 2014 to November 2015”.</li> </ul>	<p>We have incorporated your suggestion in the text.</p>
<p>“Figure 1 demonstrates the periods involved in the SSB reduction initiative”. I would say the figure does not “demonstrates the period”. Why did you choose only two centres and how did you select them? The legend cannot be</p>	<p>We believe the reviewer is referring to Figure S1, rather than Figure 1 and we apologise for the confusion. Figure 1 is located in the main text, and demonstrates the periods of the study.</p>

<p>understood from the reader: how are the centres fitted?</p> <p>How is the counterfactual estimated? There is no explanation regarding what Figure 1 shows throughout the paper. Some explanation is provided to reviewer 1- comment 4, this needs to be improved and included in the manuscript since the reader will likely not read the review document</p>	<p>Figure S1 can be found in the supplementary file and is referred to on Page 6 in the manuscript: <i>“We decided to fit individual ITSA models to overcome the large variability in volume and dollar sales data observed across centres and the different seasonal patterns depending on the type of centre (see Figure S1 in the Supplement for examples); and to account for the fact that initiative implementation was likely to differ across centres.”</i></p> <p>The detailed explanation regarding models can be found in the Analysis section of the paper (page 6-7). We have updated the footnote of Figure S1 to include reference to the model equation.</p> <p>We are able to duplicate this information in the supplement on request of the reviewer or editor, however since we have included it in the main manuscript we will leave as it is for now.</p>
<p>I believe the analysis should be re-thinked to take into account the peculiar structure and characteristics of the dataset.</p>	<p>Thank you for this comment. We appreciate that there are alternative methods of assessing food retail interventions. However, this method has been used in numerous published papers to assess real-world implementation of similar initiatives e.g. Breeze et al 2018, PLoS One. <i>The impact of local sugar sweetened beverage health promotion and price increase on sales in public leisure centre facilities.</i></p> <p>We are confident that our chosen analytical approach is appropriate to answer our research question as long as the assumptions underlying the methods are valid. To this end our models control for factors such as seasonal and centre attendance. We have aimed to provide further clarity in the methods, and discussions of the limitations.</p>

### VERSION 3 – REVIEW

<b>REVIEWER</b>	Juan Caro Health Policy Department University of North Carolina at Chapel Hill United States
<b>REVIEW RETURNED</b>	16-Nov-2019
<b>GENERAL COMMENTS</b>	I appreciate the opportunity to review this revised version. I believe comments have been partially addressed. However, some issues about identification persist. For example, observing the new presented raw-data (Figure S2), it is clear that in many centers, there was a substantial downward trend before implementation. As such, minimum assumptions for ITSA are invalid, specially for the pooled analysis. I urge the authors to address this issue, present adequate sensitivity analysis and discuss the implications.

<b>REVIEWER</b>	Beatrice Biondi University of Bologna, Italy
<b>REVIEW RETURNED</b>	28-Nov-2019

<b>GENERAL COMMENTS</b>	<p>I am overall satisfied with the changes made by the authors. I still have minor comments (see below). Moreover, I recommend an in-depth read of the paper as there could be some errors as a result of all the changes.</p> <ul style="list-style-type: none"> <li>• Pag 6: "S1, S2 and S3 are indicators of autumn (March – May), winter (June – August) and spring (September – November) respectively". Correct months.</li> <li>• Pag 9: "Fourteen of the sixteen centres had a reduced number of 'red' cold packaged drinks for sale by December 2015, however only one centre had fully implemented the initiative (i.e. did not sell a non-sports drink 'red' cold packaged drink at this time point), see Supplement Table S3." From table S3 I see 13 out of 16 centres with reduced number of 'red' cold packaged drinks, and no centres with zero red drinks available. Show average in the table. Lower and upper CI limits should be presented in this order (lower, upper). Initiative Implementation section could be improved in clarity, e.g. why is a p-value reported with no reference to any test?</li> <li>• Pag 9: " 'Green' cold packaged beverage volume sales neither changed at December 2015 (0.0% 95% CI -13.3%, 13.2%), see Figure 4, nor at December 2016 (Table S4). Stratification by variables of interest revealed observable differences in 'green' drink volume sales between centres that had removed 'red' soft drinks (9.9%, 95%CI -6.8%, 26.5%) and those that had not (-9.9%, 95%CI -30.5%, 10.7%). There were observable differences between centres with pools (13.7%, 95%CI -2.5%, 30.0%) and without pools (-17.7%, 95%CI -39.6%, 4.1%)." These observable differences are not significant.</li> <li>• Results reported in table S4 have sometimes too wide confidence interval to be meaningful, this should be acknowledged as a limitation.</li> <li>• Discussion: "These changes were generally sustained one-year post implementation, although the decline in total sales was somewhat ameliorated". The authors should acknowledge here that, based on the available data, it was not possible to evaluate the difference between the two periods, given the wide confidence interval.</li> </ul>
-------------------------	---

### VERSION 3 – AUTHOR RESPONSE

<p><b>Reviewer: 1</b>  <b>Reviewer Name:</b> Juan Caro  <b>Institution and Country:</b> Health Policy Department, University of North Carolina at Chapel Hill, United States  <b>Please state any competing interests or state 'None declared':</b> None declared  <b>Please leave your comments for the authors below</b></p>	
--	--

<p>I appreciate the opportunity to review this revised version. I believe comments have been partially addressed. However, some issues about identification persist. For example, observing the new presented raw-data (Figure S2), it is clear that in many centers, there was a substantial downward trend before implementation. As such, minimum assumptions for ITSA are invalid, specially for the pooled analysis. I urge the authors to address this issue, present adequate sensitivity analysis and discuss the implications.</p>	<p>Thank you for your comment. We have previously noted in the manuscript (page 11) that the downward trend in 'red' drink sales before 2015 is <i>not unexpected</i> due to declining soft drink consumption in Australia (Australian Bureau of Statistics). Besides, as described in the "Statistical approach and rationale" section (page 6-7) any trend occurring before implementation has been taken into account by the segmented regression approach. Our model (see Eq 1, page 6) implies a parametrization that allows for different slopes at different periods (pre-implementation, implementation, post-intervention). The main assumption of the model is that after accounting for seasonality and attendance the trend for each period is linear. Because the effect of the intervention was estimated at the centre level, the only impact that we can foresee for centres with an important decline in red drink volume during the pre-implementation period is that they will have a smaller estimated "effect" <math>\Delta_t</math>.</p> <p>We are confident that we have adequately handled the model assumptions and that our sensitivity analyses have covered a range of possible alternative scenarios.</p> <p><u>References</u></p> <p>Australian Bureau of Statistics. Children's risk factors. Canberra: Commonwealth of Australia; 2012 29/10/2012.</p>
<p><b>Reviewer: 4</b>  <b>Reviewer Name: Beatrice Biondi</b>  <b>Institution and Country: University of Bologna, Italy</b>  <b>Please state any competing interests or state 'None declared': None declared</b>  <b>Reviewer(s)' Comments to Author:</b></p>	
<p>Please leave your comments for the authors below</p> <p>I am overall satisfied with the changes made by the authors. I still have minor comments (see below). Moreover, I recommend an in-depth read of the paper as there could be some errors as a result of all the changes.</p>	<p>Thank you for your comments. We have conducted an in-depth read of the paper to pick up any errors.</p>
<ul style="list-style-type: none"> <li>• Pag 6: "S1, S2 and S3 are indicators of autumn (March – May), winter (June – August)</li> </ul>	<p>Thank you for this comment. We are unsure what is incorrect with the months. These months represent Southern hemisphere seasons.</p>



and spring (September – November) respectively”. Correct months.	
<p>Pag 9: “Fourteen of the sixteen centres had a reduced number of ‘red’ cold packaged drinks for sale by December 2015, however only one centre had fully implemented the initiative (i.e. did not sell a non-sports drink ‘red’ cold packaged drink at this time point), see Supplement Table S3.” From table S3 I see 13 out of 16 centres with reduced number of ‘red’ cold packaged drinks, and no centres with zero red drinks available.</p> <p>Show average in the table.</p> <p>Lower and upper CI limits should be presented in this order (lower, upper)</p>	<p>Thank you for this comment. We have double checked Supplement Table S3, and edited the text to reflect that 13 centres reduced the number of ‘red’ cold packaged drinks, and that no centres had completely met the initiative at December 2015.</p> <p>We have added further detail to Table S3 that recognises that the ‘red’ drinks available at December 2015 include sports drinks, which were not targeted for removal.</p> <p>We have added in the average for number of red packaged drinks available in November 2014 and December 2015 in Table S3.</p> <p>We have corrected the upper and lower limits in Table S3.</p>
Initiative Implementation section could be improved in clarity, e.g. why is a p-value reported with no reference to any test?	<p>We have added further detail to the methods section, page 5, rather than the discussion.</p> <p><i>“Paired t-tests were used to determine whether the difference between the number of ‘red’, ‘amber’ and ‘green’ drinks available for purchase between these two timepoints was statistically significant”.</i></p>
<p>Pag 9: “ ‘Green’ cold packaged beverage volume sales neither changed at December 2015 (0.0% 95% CI -13.3%, 13.2%), see Figure 4, nor at December 2016 (Table S4). Stratification by variables of interest revealed observable differences in ‘green’ drink volume sales between centres that had removed ‘red’ soft drinks (9.9%, 95%CI -6.8%, 26.5%) and those that had not (-9.9%, 95%CI -30.5%, 10.7%). There were observable differences between centres with pools (13.7%, 95%CI -2.5%, 30.0%) and without pools (-17.7%, 95%CI -39.6%, 4.1%).”</p> <p>These observable differences are not significant.</p>	<p>Thank you for this comment, we have edited the text to indicate that these results are non-significant. Page 9</p>
Results reported in table S4 have sometimes too wide confidence interval to be meaningful, this should be acknowledged as a limitation.	<p>We have added this to the limitations, page 11</p> <p><i>Wide confidence intervals for some outcomes at December 2016 (i.e. ‘green’ drink volume) limit the interpretability of the results.</i></p>

Discussion: "These changes were generally sustained one-year post implementation, although the decline in total sales was somewhat ameliorated". The authors should acknowledge here that, based on the available data, it was not possible to evaluate the difference between the two periods, given the wide confidence interval.	Thank you for this comment. We have changed the text to reflect this page 11:  <i>"The decline in total sales was somewhat ameliorated at one year, however such evidence is inconclusive due to the wide confidence intervals."</i>
---	--

#### VERSION 4 – REVIEW

<b>REVIEWER</b>	Juan Caro Health Policy and Management Department, University of North Carolina at Chapel Hill, USA
<b>REVIEW RETURNED</b>	11-Dec-2019

<b>GENERAL COMMENTS</b>	As the authors have not addressed the statistical validity of their results with adequate sensitivity analysis. Moreover, in the light of the data, the methods are not valid to achieve any type of causal inference from the implemented intervention.
-------------------------	--

<b>REVIEWER</b>	Beatrice Biondi University of Bologna, Italy
<b>REVIEW RETURNED</b>	13-Dec-2019

<b>GENERAL COMMENTS</b>	I believe the authors replies and changes solved my concerns.
-------------------------	---

#### VERSION 4 – AUTHOR RESPONSE

##### Response to Reviewer 1

Thank you for your comments. We have responded to each of your main concerns below.

**Analysis method:** As we have previously argued, the downward trend in red drink purchases prior to policy implementation is adjusted for using the ITSA approach. Indeed, a key strength of the ITSA approach is that it takes into account pre-implementation trends. The approach we have used is valid, and has been used in numerous other public health policy evaluations. We believe that the use of analysis method is at this point, a matter of opinion.

**Sensitivity analysis:** The sensitivity analyses that we have presented have covered a range of possible alternative scenarios. Unfortunately, we would require more specificity in order to provide the additional sensitivity analysis that you are requesting. For example, you refer to "identification issues" however it is not clear what parameters you are referring to that cannot be identified under our approach.

Consequently, we have been unable to make revisions in response to your comments.

##### Response to Reviewer 4

Reviewer 4 has recommended no changes.