General comment
The Rehm et al paper provides a useful contribution to the increasing number of studies aimed at evaluating the impact of MUP on a range of outcomes, including alcohol consumption. The authors supplement their earlier work using the Kantar Worldpanel household panel data by using Kantar’s repeat cross-sectional drinking diary dataset, Alcovision. This dataset has been used previously to better understand drinking occasions (Ally et al, 2016) and is planned for use in a MUP evaluation study being conducted as part of the comprehensive evaluation being led by Public Health Scotland (see here). The value of the dataset lies not only in its frequency of data collection, but also its comparatively large sample size (check) and the fact that consumption estimates can be broken down by sociodemographic characteristics, thereby providing the opportunity to explore the differential impact of MUP. This makes it more useful for evaluation purposes than national annual surveys across GB. It is recommended that the strength of the survey is more clearly described in the Introduction.

Consumption level
Assessing differential impact is important because evidence in support of MUP has emphasised the fact that it is a policy that is likely to disproportionately impact those at highest risk of alcohol-related harms: men, those in more deprived areas, and the heaviest drinkers. It should be noted, however, that as a policy, MUP was implemented to target hazardous and harmful drinkers, not any particular sex or age group (https://www.gov.scot/publications/minimum-unit-pricing-alcohol-final-business-regulatory-impact-assessment/). A major limitation, in this study, therefore, is the lack of analyses by consumption level, as has been done in the research team’s prior research. It is those who drink the most amount of alcohol who are most likely to purchase alcohol at the lower end of the price spectrum. It is therefore advised that this analysis is undertaken.
Analytical method
The study uses controlled interrupted time series analysis as its primary analytical method (as indicated in the title). This is an appropriate approach for evaluating 'natural experiments', particularly when dealing with high frequency data. However, the study also supplements this with a before-and-after analyses, which doesn’t have prominence in the abstract or title, but does in the results and discussion. The authors should be much clearer about the rationale for the before and after study and, in particular, the rationale for its inclusion. I wonder if the unusually hot summer in 2018 alongside the short follow-up period have particular implications for this analysis, which only includes country as a factor, and which doesn’t not take into account underlying seasonal trends?

Limitations
Related to the above, I think the authors could expand much more on the limitations of the study in relation to the data (e.g. quota sampling; sampling bias) and the analytical method (particularly the before- and after analysis), and consider how this might affect their interpretation of results.

Explaining the results
The authors should seek to explain the results that they have observed and the conclusions that they have made. For example, why would MUP be associated with increased consumption among particular groups, if not explained by study limitations?

Specific comments

<table>
<thead>
<tr>
<th>Page 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 5: The objective could be worded more clearly. For example: To assess if the impact on alcohol consumption of introducing MUP for alcohol in Scotland differs between men and women.</td>
</tr>
<tr>
<td>Line 11: Consider if the name of the survey represents</td>
</tr>
<tr>
<td>Line 16: Suggest replacing ‘price’ with ‘pricing’ (applies in other places too)</td>
</tr>
<tr>
<td>Line 21 (and elsewhere): I would strongly discourage an emphasis on whether results were statistically significant or not and instead focus on the size of the effect estimate and the uncertainty around it. Phrases such as almost significant should be completely avoided. For justification see here:</td>
</tr>
<tr>
<td>Line 34: “that there are no consistent greater changes with greater deprivation potentially” This is inconsistent with the Discussion which states that “For the sample as a whole, those who lived in less deprived areas showed the greatest reduction in consumption.” It also seems that this may be drawn from the before-after analyses which isn’t mentioned in the abstract.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 23: I suggest you refer to consumption effects (either instead of or in addition to mortality) as this is the focus of the paper</td>
</tr>
<tr>
<td>Line 29: Ref 13 is an inappropriate reference for this statement. The blog piece does not represent an evaluation of the policy. In addition, mortality data area available by sex so this should be possible.</td>
</tr>
<tr>
<td>Line 31: the underlying assumption of the policy is that it would target hazardous and harmful drinkers.</td>
</tr>
<tr>
<td>Line 10: Given the dataset has an individual measure of socioeconomic position (social class), please explain why an arguably less sensitive area-based measure has been used.</td>
</tr>
</tbody>
</table>
Line 23: Please consider if the use of country-specific IMDs has implications for the calculation of net consumption (i.e. Scotland minus England) given the deprivation profile within countries is likely to be different (e.g. the most deprived fifth of areas in Scotland are more deprived than the most deprived fifth in England). Would a standardised measure of deprivation be more robust?

Line 47: Given the relatively small sample size in some subgroups, it would be useful to know what size of effect the sample size was powered to detect.

Page 5
Line 27: Please provide further justification for the before and after analyses and how it enables a better understanding of the variation in the impact of MUP. Is this sensitivity analyses or supplementary analyses or part of the main analyses?

Page 6
Line 21: The consumption decline with age is inconsistent with patterns seen in national population surveys. This might be a point worth noting in the discussion.

Line 26 and elsewhere: CI should be in grams (or other relevant units e.g. %)

Page 7
The authors should provide a brief summary of the descriptive analyses shown in the appendix. Also what analytical method was used to examine the association between age/deprivation and consumption; the patterns are clearly not linear

Table 1: It is useful to have the detailed outputs from the model; However, please consider alternative approaches to presenting the data as the main effect estimate of the primary outcome is not prominent.

Table 1: Please remove the shading denoting the arbitrary statistical significance threshold.

Figure 1: Please supplement the figures in the paper with tabulation of all results in an appendix.

Line 48: I wonder to what extent the hot summer (and football world cup) plays a role in the increase in consumption among young people. I realise this is based on controlled analysis, but the weather was so atypical and is such a driver of consumption, it should be discussed. Similarly, other potential explanation should be given – young people spend more on alcohol? Cross-border purchasing?

Line 50: The confidence intervals for IMD groups are large; was testing done to assess differences between IMD groups (e.g. IMD3 versus the rest?). It is also worth considering what might explain this finding in the middle deprivation group and not the others.

Page 8
Line 10 and 21: I’m not sure how useful Figures 3 and 4 are given they are so heavily confounded by age/deprivation.

Line 14: As noted above, could this be explained by the hot summer?

Line 45: I’m not clear on what Figure 5 is showing and why z-scores have been used. The figure seems to suggest that the impact of MUP on alcohol consumption in quintiles 1-4 is pretty much of the same magnitude, irrespective of age? Please explain the method and the chart more clearly.

Figure 5: The y axis label should be a legend.
Line 6: Suggest the first paragraph is focuses on the results observed in the study rather than contextualising them against prior findings.

Line 7: Please consider if this comparison is appropriate given the differences in datasets; time frames etc.

Line 13: This conclusion doesn't seem justified based on the results. The interrupted time series analyses showed no obvious patterning and it could be argued that the before-after analyses showed the same (Figure 5 (which needs confidence intervals))

Line 36: This is repeated from above

Line 40 onwards: The authors should attempt to explain why this might be the case, taking into account the study's limitations and the findings from other studies.

Page 10
It would be useful to know the purpose of the funds to fully understand the nature of the competing interest.

Page 14
Figure 4: Please clarify if it is 'controlling for changes in England' in this case, or 'adjusting for age'

Page 19
If retaining these figures please put them on the same scale to enable read across.

Page 28
Table 1: Please also show the sample size for age groups and deprivation groups (as analyses was also done at this level).

Page 33
Item 1: No mention of before-after
Item 2: Deprivation should also be mentioned.
Item 43: Sensitivity analyses should always be considered when conducting observational studies. Consider how interpretation of your results might be strengthened by, for example, examining the difference between Scotland and Northern England, which has a more similar demographic and alcohol consumption/harm profile than England as a whole.

---

**REVIEWER**
Norberto Francisco Hernandez-Llanes
Secretaria de Salud de Mexico, Comision Nacional contra las Adicciones

**REVIEW RETURNED**
29-Jun-2021

**GENERAL COMMENTS**
I am very honored to have reviewed the manuscript "Differential impact of minimum unit pricing on alcohol consumption between Scottish men and women: controlled interrupted time-series analysis". My sincere appreciation to the authors, for the work they are presenting.

I have a couple of comments that I do not consider to affect the conclusions of the manuscript, and that I politely bring to the authors' consideration:

1. I understand that the studied sample of the Kantor Panel was chosen to be representative of the general population of the United Kingdom. However, given that the results point to a greater effect in the population with less deprivation, my question is whether the authors considered an effect of underrepresentation of people with greater deprivation (and therefore with less accessibility to the internet), or if the Internet accessibility does not represent a problem in the weighting of the sample studied.
2. I leave it to the authors’ judgment whether Figure 5 is necessary or can be sent to the appendix, since I consider that the information in the text is sufficient to explain this relationship.

REVIEWER
Mark Meyer
Georgetown University, Mathematics and Statistics

REVIEW RETURNED
17-Sep-2021

GENERAL COMMENTS
Statistical Review for BMJ Open manuscript bmjopen-2021-054161
“Differential impact of minimum unit pricing on alcohol consumption between Scottish men and women: controlled interrupted time-series analysis.”

The authors implement an interrupted time series regression model with control to assess the impact of minimum unit pricing on alcohol in consumption in Scotland. England serves as an additional control and the authors aim to determine the overall impact as well as the sex-specific impact. I have several concerns with both the design of the study and the implementation of the statistical models. To summarize quickly the major comments that are below, I believe the authors could benefit from consultation with and support from a biostatistician trained in longitudinal modeling. Major comments follow immediately with Minor comments after:

1. The authors’ implement their models using SPSS and should provide syntax files for replication or so the reader can at least see exactly what SPSS is running.

2. The manuscript contains a large number of technical terms related to the substantive area that need to be defined. BMJ Open is read by a wide range of readers who may not be familiar with terms such as ‘on-premise’ or ‘on-trade consumption.’ These terms should be briefly defined in the manuscript when introduced.

3. The authors need to describe in greater detail how the sample was obtained and what population the sample is taken from. Was the sample random, stratified or otherwise? How were potential participants identified? The authors mention ‘quota samples’ which I assume means some form of stratified sampled. Does the resulting sample match what is expected in each population? The authors appear to take no steps to ensure the sample is representative of the population upon which they are drawing inference.

4. The statistical modeling for the main result needs more careful treatment. Longitudinal models are capable of addressing some of what the authors attempt to ‘test away,’ for example potential auto-correlation within sample. Weekly averaging of responses to survey acts as a coarse smoother, un-intentionally denoising the signal. One imagines variation in consumption within week is high. The authors appear to implement a multiple regression for the main result. Post-stratification is mentioned in the discussion but does not appear to be implemented if the sample is not population representive, this potential omission is problematic. The analysis as it stands makes some interesting assumptions: by pre-determining the authors assume that weeks in England are matched to weeks in Scotland by their calendar alignment. England and Scotland are also treated as large regions. One imagines regional differences play a role (hence the inclusion of this deprivation variable). Mixed effects models could help model a sub-regions latent tendency to respond...
within each nation while still achieving the over-all inference goal of assessing the impact of minimum unit pricing using England as a control|here a "di erence-in-di erence" model would result for the overall model. In other words, compare Scot-land’s pre- vs post-consumption to England’s pre- vs post-consumption while controlling for sub-region clustering.

5. England’s use as a control makes sense given its proximity to Scotland and, presumably, similar behavior in terms of consumption. But the authors need to better establish the control beyond the visual inspection of weekly averages. Demographic comparison, for one, to ensure the nations have similar population characteristics would be useful. Even a simple test for parallelism pre-MUP implementation would strengthen claims of similar consumption. That said, ITS provides a control with itself in the pre-vs post-analysis. The use of an external control is a nice touch, but care must be taken to demonstrate that the populations are indeed comparable in every way except for the intervention. Using Scotland as its own control is, in some ways, a stronger justification than saying England is nearby and has visually similar consumption. More justification is therefore needed.

6. The authors perform a range of what I’d call "sub-group" analyses with analyses performed by age group, deprivation group, sex, etc. Beyond the sex breakdown in Table 1 of the appendix, I’m not seeing a discussion of the number of subjects available for analysis in each of these sub-groups. Sub-group analyses can be useful to identify, as the authors do, the drivers of the overall results. But care must be taken to ensure these are correctly powered (and they likely are, but the proof of this is missing). A traditional "Table 1" would be useful to provide both information about the subgroups and allow the authors to demonstrate whether the sample is representative and the comparability of the two nations.

7. The use of negative binomial model is incorrect for the authors’ data type. Negative binomial modeling does address overdispersion, but for overdispersed count data where the Poisson assumption of equality between the mean and variance is violated. In other words, the model requires a discrete and specifically count outcome|the negative binomial distribution is a discrete distribution. The authors’ outcome is not a count but an average of consumption measured in grams. This is a continuous random variable and should be treated as such. The authors could implement a transformation to address the skewness in the data or build a model that allows for skewness in a continuous outcome, like a log-normal regression or gamma regression model.

8. The authors should explain further why it is the case that "there is no reason to believe that under-reporting would di er by sex." This is particularly pertinent when using data from a survey that the authors appear to have had no control over the design of.

Minor comments:
1. The manuscript as a whole should be thoroughly proof-read for errors and awkwardly worded sentences. For example, the rst sentence of the ‘Study design’ subsection in Methods needs to be reworked.

2. What is the reason a weekly average is used? Some justification of this would be useful; it could include a desire to smooth out variation if the within variability is quite high, though reducing variability before analyzing can have issues.

3. Why are the ITS models di erenced between country rst but the before-and-after analysis not?

4. For the sub-group analyses, the authors standardize the outcomes to assure comparability of model coe cients. A joint modeling approach where age group is a predictor could be used without standard-ization. Formal tests can then be constructed.

5. Some of what appear to be the ITS sub-group results are under the heading ‘before-and-after analysis’ despite not being characterized as such in the statistical analysis section.

6. The abbreviation MUP should be de ned parenthetically or otherwise in the abstract after minimum unit pricing.

7. Is there evidence of cross-consumption? Have people in Scotland living near the border began consuming in Northern England as a result of the MUP implementation? The analysis as it stands assumes there is no crossover between Scotland and the control.

8. The gures on pages 18 and 19 are too small and, consequently, hard to read.

---

**REVIEWER**
Jim Lewsey  
University of Glasgow

---

**GENERAL COMMENTS**
I congratulate the authors on a very interesting paper on an important topic. It is well written and is well presented in the main. Controlling the analyses for England is a key strength. However, I have concerns that at present the analyses presented do not provide statistical evidence to support the interpretations you are making about differential MUP effects. Further, I think more reassurance on possible selection/sampling biases not being a major threat to internal validity is required.

**Major:**
- What is the response rate for KWP? Is it similar between Scotland and England? It is stated that residents from Scotland and 18-34 year olds are oversampled but looking at appendix Figs 1&2 the oversampling looks similar in England and Scotland? Further, it looks like the difference with total population %s is more pronounced for women than men. Given this study is looking at differences between men and women isn’t this a big threat to the internal validity of the study?
- The ITS statistical analyses undertaken do not formally test whether any MUP effect is different between men and women. It is possible to set up the data set as a panel and then use interaction...
terms in the regression to statistically test for differential effects by sex (and deprivation). I recommend the authors do this (or provide a rationale for why this was not done). Of note, this type of approach seems to have been taken for the ‘before and after analyses’ (models 1&2 include interactions involving MUP&age and MUP&deprivation. Again though, why run separate models for men and women given the aim of this study?

- Results, ITS: claims such as “Associated significant differences in consumption were restricted to women…” are only supported with descriptive interpretations of separate model results for men and women. Interaction tests (see previous point) are needed to support these claims. Same for the paragraph interpreting results shown in Figs 1&2.
- Results, before and after analyses: some interaction tests were done here but are not reported on. However, separate models were done for men and women so again, at present, the associated interpretation is descriptive and not statistically tested for.

Minor:
- Abstract: Key words; is it reasonable to have ‘health inequalities’ when the outcome measure in this study is alcohol consumption (not directly health).
- The labelling and colour scheme in Figs 1&2 of the Appendix make it difficult to know what each line refers to. Can this be made more clear?
- What was the rationale for grouping age in the categories you chose?
- “We adjusted the dependent variables for any seasonality” – how exactly was this done?
- In the ‘before and after analyses’, no explanation is given for why interaction terms are added to models’ 1&2 (and then the results of these are not interpreted).

Typos, etc.:
- Abstract: “Although drops in consumption varied by level of residential deprivation, <and differently for men and women>…”; missing word(s)? Sentence does not quite ‘scan’
- Appendix: “Deprivation” instead of “Deprivation”
RESPONSE: Thank you for all your very helpful comments. We have added a sentence on the strength of the Alcovision survey to the introduction (see page 4).

Consumption level
Assessing differential impact is important because evidence in support of MUP has emphasised the fact that it is a policy that is likely to disproportionately impact those at highest risk of alcohol-related harms: men, those in more deprived areas, and the heaviest drinkers. It should be noted, however, that as a policy, MUP was implemented to target hazardous and harmful drinkers, not any particular sex or age group (https://www.gov.scot/publications/minimum-unit-pricing-alcohol-final-business-regulatory-impact-assessment/). A major limitation, in this study, therefore, is the lack of analyses by consumption level, as has been done in the research team’s prior research. It is those who drink the most amount of alcohol who are most likely to purchase alcohol at the lower end of the price spectrum. It is therefore advised that this analysis is undertaken.

RESPONSE: We have added this analysis by consumption level, finding in general that heavier drinkers reduced their alcohol consumption more than lighter drinkers, with the exception of the top 5% of heaviest drinking men who seemed to increase their consumption associated with the introduction of MUP.

Analytical method
The study uses controlled interrupted time series analysis as its primary analytical method (as indicated in the title). This is an appropriate approach for evaluating ‘natural experiments’, particularly when dealing with high frequency data. However, the study also supplements this with a before-and-after analyses, which doesn’t have prominence in the abstract or title, but does in the results and discussion. The authors should be much clearer about the rationale for the before and after study and, in particular, the rationale for its inclusion. I wonder if the unusually hot summer in 2018 alongside the short follow-up period have particular implications for this analysis, which only includes country as a factor, and which doesn’t take into account underlying seasonal trends?

RESPONSE: We have added a rationale and better explanation for the before- and after analyses. We undertake it as a validity check to better understand variation in the associated impact of MUP by age and deprivation for each individual age and each individual deprivation score, rather than by the four age groups and the five deprivation groups used in the interrupted time series analyses. As explained in the methods (and revised now to explain more clearly), each individual age and each individual deprivation score (as an integer) are added to the model as DUMMY VARIABLES, and NOT COVARIATES. In the model, we include an interaction term that is country*event*age (or deprivation score), where event is the introduction of MUP. From this, we look at the differences in the coefficients (Scotland*event*age (or deprivation score)) minus (England*event*age (or deprivation score)); these differences tell us the associated impact of MUP in Scotland for each individual age (or deprivation score), controlling for any changes in England from before to after the introduction of MUP for that particular age (or deprivation score); this should take into account common occurrences (such as a hot summer) that occurred in both Scotland and England. The follow-up period included data from 17596 respondents. The findings of the before and after analyses are not inconsistent with the grouped findings from the interrupted time series analyses.

Limitations
Related to the above, I think the authors could expand much more on the limitations of the study in relation to the data (e.g. quota sampling; sampling bias) and the analytical method (particularly the before- and after analysis), and consider how this might affect their interpretation of results.

RESPONSE: The limitations with relation to the data are described to a considerable extent in the discussion section. Interrupted time series analyses with a location control is the most appropriate method to analyse a natural experiment, such as the introduction of MUP, and we have followed the guidance in the field of addictions (e.g., reference 22 in the paper’s references). With undertaking interrupted time series analyses, it is also not uncommon to include before and after analyses, recognizing that such analyses do not take into account such issues as seasonality. We have added a sentence of such limitations in the discussion section (see page 17).
Explaining the results
The authors should seek to explain the results that they have observed and the conclusions that they have made. For example, why would MUP be associated with increased consumption among particular groups, if not explained by study limitations?

RESPONSE: Response variability is expected, especially as MUP is a population level intervention. Indeed, individual variability has been anticipated following MUP in Scotland and this would tend to be in heavy/dependent drinkers who are predominantly men. We have added text on page 17 to discuss this and also added references 29 and 30 in the text.

Specific comments

Page 3
Line 5: The objective could be worded more clearly. For example: To assess if the impact on alcohol consumption of introducing MUP for alcohol in Scotland differs between men and women.

RESPONSE: We have edited the abstract and have included: “We analysed consumption data of 106,490 adults to assess the immediate impact of MUP, and whether the impact differed by sex, level of consumption, age, social grade, and level of residential deprivation.”

Line 11: Consider if the name of the survey represents
Line 16: Suggest replacing ‘price’ with ‘pricing’ (applies in other places too)

RESPONSE: The abstract has been edited.

Line 21 (and elsewhere): I would strongly discourage an emphasis on whether results were statistically significant or not and instead focus on the size of the effect estimate and the uncertainty around it. Phrases such as almost significant should be completely avoided. For justification see here:

RESPONSE: Thank you. We have checked and adjusted throughout, focusing on the size of the effect and the uncertainty surrounding it.

Line 34: “that there are no consistent greater changes with greater deprivation potentially” This is inconsistent with the Discussion which states that “For the sample as a whole, those who lived in less deprived areas showed the greatest reduction in consumption.” It also seems that this may be drawn from the before-after analyses which isn’t mentioned in the abstract.

RESPONSE: We have edited the abstract and the summary box to ensure consistency throughout the paper.

Page 4
Line 23: I suggest you refer to consumption effects (either instead of or in addition to mortality) as this is the focus of the paper
Line 29: Ref 13 is an inappropriate reference for this statement. The blog piece does not represent an evaluation of the policy. In addition, mortality data area available by sex so this should be possible.

RESPONSE: Yes, we have adjusted the text and deleted reference 13.

Line 31: the underlying assumption of the policy is that it would target hazardous and harmful drinkers.

RESPONSE: Thank you - noted and adjusted text.
Line 10: Given the dataset has an individual measure of socioeconomic position (social class), please explain why an arguably less sensitive area-based measure has been used.

**RESPONSE:** We have added new analyses also by social grade. The advantage of area-based residential deprivation is that it is based on multiple indices and provides a continuous variable rather than a grouping of 4 (AB, C1, C2, DE). In the discussion, we have noted (page 16) that the risk of alcohol-related harm increasing both the more socio-economically disadvantaged the individual is, and, over and above that, the more socially disadvantaged the residential area in which the individual resides.

Line 23: Please consider if the use of country-specific IMDs has implications for the calculation of net consumption (i.e. Scotland minus England) given the deprivation profile within countries is likely to be different (e.g. the most deprived fifth of areas in Scotland are more deprived than the most deprived fifth in England). Would a standardised measure of deprivation be more robust?

**RESPONSE:** In terms of absolute deprivation, this might be correct; in terms of relative deprivation, we compare like with like - i.e., the most deprived fifth with the most derived fifth. We have added a sentence about this on page 16 in the discussion.

Line 47: Given the relatively small sample size in some subgroups, it would be useful to know what size of effect the sample size was powered to detect.

**RESPONSE:** We have added a paragraph on Power considerations to the Methodology section.

Page 5
Line 27: Please provide further justification for the before and after analyses and how it enables a better understanding of the variation in the impact of MUP. Is this sensitivity analyses or supplementary analyses or part of the main analyses?

**RESPONSE:** We have added a rationale and better explanation for the before- and after analyses. We undertake it as a validity check to better understand variation in the associated impact of MUP by age and deprivation for each individual age and each individual deprivation score, rounded to an integer, rather than by the four age groups and the five deprivation groups used in the interrupted time series analyses. This gives us more detailed information about the relationships between both age and deprivation and the potential impact of MUP. Within the before and after analysis, we have also added a further sensitivity analysis that is consistent with the main findings of the before and after analysis.

Page 6
Line 21: The consumption decline with age is inconsistent with patterns seen in national population surveys. This might be a point worth noting in the discussion.

**RESPONSE:** As this is not a focus of the study, we have not highlighted the issue in the discussion.

Line 26 and elsewhere: CI should be in grams (or other relevant units e.g. %)

**RESPONSE:** We have clarified the units throughout.

Page 7
The authors should provide a brief summary of the descriptive analyses shown in the appendix. Also what analytical method was used to examine the association between age/deprivation and consumption; the patterns are clearly not linear. Table 1: It is useful to have the detailed outputs from...
the model; However, please consider alternative approaches to presenting the data as the main effect estimate of the primary outcome is not prominent.

**RESPONSE:** We have moved reference to the figure plotting mean deprivation score by age to the methods section, as we think it better placed there. We have described the J-shaped pattern in the text. The graph is simply a plot of mean deprivation score by age – this is stated in the revised legend to the graph, Appendix Figure 7, page 4.

We have added a new Figure 1 plotting the changes over time that illustrates the main findings. We have left Table 1 as it is. We think that the revised paragraph introducing the table is clear enough in emphasizing the main findings.

With respect to the relationship between consumption and age and deprivation score of the respondents, these have been added as graphs to the Appendix. We used age and deprivation score in the before and after analyses. But, here, as independent variables, age and deprivation score (rounded to an integer) were entered as dummy-coded variables for each age and each deprivation score, not as continuous variables (see methods section).

Table 1: Please remove the shading denoting the arbitrary statistical significance threshold.

**RESPONSE:** Shading removed.

Figure 1: Please supplement the figures in the paper with tabulation of all results in an appendix.

**RESPONSE:** We have added tables of the data to the appendix.

Line 48: I wonder to what extent the hot summer (and football world cup) plays a role in the increase in consumption among young people. I realise this is based on controlled analysis, but the weather was so atypical and is such a driver of consumption, it should be discussed. Similarly, other potential explanation should be given – young people spend more on alcohol? Cross-border purchasing?

**RESPONSE:** We do not know why this might be, and, given the data we have, it is difficult to speculate. We have added some text to the discussion (see page 17), also calling for further analyses. For the hot summer our analyses focussed on the differences between Scotland and England, this would mean that the hot summer was differently affecting Scotland vs. England/North England, which seems implausible.

Line 50: The confidence intervals for IMD groups are large; was testing done to assess differences between IMD groups (e.g. IMD3 versus the rest?). It is also worth considering what might explain this finding in the middle deprivation group and not the others.

**RESPONSE:** When undertaking the new analyses by social grade, we noticed a data error in the data set used for analysis of the deprivation groupings. When correcting the error, the IMD3 discrepancies disappeared. We double checked all other data sets used for the analyses and found no other errors. Thus, all other findings remain the same.

Page 8
Line 10 and 21: I’m not sure how useful Figures 3 and 4 are given they are so heavily confounded by age/deprivation.

**RESPONSE:** Now Figures 4 and 5 are the main findings of the before and after analysis, and we consider that they should be left in. We demonstrate that the slopes by age of Figure 4 do not change by deprivation group.
Line 14: As noted above, could this be explained by the hot summer?

**RESPONSE:** See response above.

Line 45: I’m not clear on what Figure 5 is showing and why z-scores have been used. The figure seems to suggest that the impact of MUP on alcohol consumption in quintiles 1-4 is pretty much of the same magnitude, irrespective of age? Please explain the method and the chart more clearly.

**RESPONSE:** See response above. We have deleted Figure 5 and added text stating that the slopes by age of now Figure 2 do not differ by deprivation group.

Figure 5: The y axis label should be a legend.

**RESPONSE:** Figure 5 now deleted.

Page 9
Line 6: Suggest the first paragraph is focussed on the results observed in the study rather than contextualising them against prior findings.

**RESPONSE:** We have structured the discussion to start with the findings of the present study, and then compare them to prior findings of other studies.

Line 7: Please consider if this comparison is appropriate given the differences in datasets; time frames etc

**RESPONSE:** We have revised the discussion, and we think it clearer now and that the conclusions are justified based on the results.

Page 10
It would be useful to know the purpose of the funds to fully understand the nature of the competing interest.

**RESPONSE:** The purpose added.

Page 14
Figure 4: Please clarify if it is ‘controlling for changes in England’ in this case, or ‘adjusting for age’

**RESPONSE:** Figure 4 in the Appendix is neither controlled for age nor for England.

Page 19
If retaining these figures please put them on the same scale to enable read across.

**RESPONSE:** Vertical y-axes for all such figures have been put on the same scale.

Page 28
Table 1: Please also show the sample size for age groups and deprivation groups (as analyses was also done at this level).

**RESPONSE:** Data added in what is now Appendix Table 2.
Item 1: No mention of before-after

RESPONSE: Figure now deleted.

Item 2: Deprivation should also be mentioned.

RESPONSE: Figure now deleted.

Item 43: Sensitivity analyses should always be considered when conducting observational studies. Consider how interpretation of your results might be strengthened by, for example, examining the difference between Scotland and Northern England, which has a more similar demographic and alcohol consumption/harm profile than England as a whole.

RESPONSE: Reported sensitivity analysis with Northern England as control. The main findings are very similar.

Reviewer: 2

Dr. Norberto Francisco Hernandez-Llanes, Secretaria de Salud de Mexico Comments to the Author:
I am very honored to have reviewed the manuscript "Differential impact of minimum unit pricing on alcohol consumption between Scottish men and women: controlled interrupted time-series analysis". My sincere appreciation to the authors, for the work they are presenting.

RESPONSE: Thank you for your helpful comments.

I have a couple of comments that I do not consider to affect the conclusions of the manuscript, and that I politely bring to the authors' consideration:

1.- I understand that the studied sample of the Kantor Panel was chosen to be representative of the general population of the United Kingdom. However, given that the results point to a greater effect in the population with less deprivation, my question is whether the authors considered an effect of underrepresentation of people with greater deprivation (and therefore with less accessibility to the internet), or if the Internet accessibility does not represent a problem in the weighting of the sample studied.

RESPONSE: Appendix figure 3 does not suggest underrepresentation. We have added this point to the discussion.

2.- I leave it to the authors' judgment whether Figure 5 is necessary or can be sent to the appendix, since I consider that the information in the text is sufficient to explain this relationship.

RESPONSE: The figure has been deleted and replaced with simplified text.
Reviewer: 3
Dr. Mark Meyer, Georgetown University

Comments to the Author:
Please see my attached comments.

The authors implement an interrupted time series regression model with control to assess the impact of minimum unit pricing on alcohol consumption in Scotland. England serves as an additional control and the authors aim to determine the overall impact as well as the sex-specific impact. I have several concerns with both the design of the study and the implementation of the statistical models. To summarize quickly the major comments that are below, I believe the authors could benefit from consultation with and support from a biostatistician trained in longitudinal modeling. Major comments follow immediately with Minor comments after:

RESPONSE: Thank you very much for all of your helpful comments.

1. The authors’ implement their models using SPSS and should provide syntax for replication or so the reader can at least see exactly what SPSS is running.

RESPONSE: We have added the syntax in boxes in the appendix for the main hypotheses.

2. The manuscript contains a large number of technical terms related to the substantive area that need to be defined. BMJ Open is read by a wide range of readers who may not be familiar with terms such as ‘on-premise’ or ‘off-premise,’ ‘residential deprivation,’ and ‘on-trade’ or ‘off-trade consumption.’ These terms should be briefly defined in the manuscript when introduced.

RESPONSE: We have illustrated or defined the terms.

3. The authors need to describe in greater detail how the sample was obtained and what population the sample is taken from. Was the sample random, stratified or otherwise? How were potential participants identified? The authors mention ‘quota samples’ which I assume means some form of stratified sample. Does the resulting sample match what is expected in each population? The authors appear to take no steps to ensure the sample is representative of the population upon which they are drawing inference.

RESPONSE: We have stated both in the methods and in the discussion that this is not a random sample. It is a quota sample where weighting is used to model the population. Annexes Figures 1-3 compare the sample to the general population for sex, age and deprivation. We refer to this again in the discussion.

4. The statistical modeling for the main result needs more careful treatment. Longitudinal models are capable of addressing some of what the authors attempt to “test away,” for example potential autocorrelation within sample. Weekly averaging of responses to survey acts as a coarse smoother, unintentionally denoising the signal. One imagines variation in consumption within week is high. The authors appear to implement a multiple regression for the main result. Post-stratification is mentioned in the discussion but does not appear to be implemented, if the sample is not pop-representative, this potential omission is problematic. The analysis as it stands makes some interesting assumptions: by pre-differencing the authors assume that weeks in England are matched to weeks in Scotland by their calendar alignment. England and Scotland are also treated as large regions. One imagines regional differences play a role (hence the inclusion of this deprivation variable). Mixed effects models could help model a sub-regions latent tendency to respond within each nation while still achieving the overall inference goal of assessing the impact of minimum unit pricing using England as a control. Here a “difference-in-difference” model would result for the overall model. In other words, compare Scotland’s pre- vs post-consumption to England’s pre- vs post-consumption while controlling for sub-region clustering.

RESPONSE: Thank you and noted. We have followed similar methodologies to previous publications [O'Donnell A, Anderson P, Jané-Llopis E, et al. Immediate impact of minimum unit pricing on alcohol consumption in Scotland. IJERPH 2021;18(21):10619]
5. England's use as a control makes sense given its proximity to Scotland and, presumably, similar behavior in terms of consumption. But the authors need to better establish the control beyond the visual inspection of weekly averages. Demographic comparison, for one, to ensure the nations have similar population characteristics would be useful. Even a simple test for parallelism pre-MUP implementation would strengthen claims of similar consumption. That said, ITS provides a control with itself in the pre- vs post-analysis. The use of an external control is a nice touch, but care must be taken to demonstrate that the populations are indeed comparable in every way except for the intervention. Using Scotland as its own control is, in some ways, a stronger justification than saying England is nearby and has visually similar consumption. More justification is therefore needed.

RESPONSE: Noted. We have added tables and figures to the appendix showing the demographic comparisons, which are similar. We have tested for and confirmed parallelism prior to the introduction of MUP (Appendix table 1). We have undertaken an additional sensitivity analysis using Northern England as a control.

6. The authors perform a range of what I'd call "sub-group" analyses with analyses performed by age group, deprivation group, sex, etc. Beyond the sex breakdown in Table 1 of the appendix, I'm not seeing a discussion of the number of subjects available for analysis in each of these sub-groups. Sub-group analyses can be useful to identify, as the authors do, the drivers of the overall results. But care must be taken to ensure these are correctly powered (and they likely are, but the proof of this is missing). A traditional "Table 1" would be useful to provide both information about the subgroups and allow the authors to demonstrate whether the sample is representative and the comparability of the two nations.

RESPONSE: Table of numbers in sub-groups added, Appendix Table 2. We have added a short explanation of the power of the study to the methods section.

7. The use of negative binomial model is incorrect for the authors' data type. Negative binomial modeling does address overdispersion, but for over dispersed count data where the Poisson assumption of equality between the mean and variance is violated. In other words, the model requires a discrete and specifically count outcome, the negative binomial distribution is a discrete distribution. The authors' outcome is not a count but an average of consumption measured in grams. This is a continuous random variable and should be treated as such. The authors could implement a transformation to address the skewness in the data or build a model that allows for skewness in a continuous outcome, like a log-normal regression or gamma regression model.

RESPONSE: For the before and after analyses, we have treated consumption (rounded to an integer) as a count variable (counting grams of alcohol consumed). We consider that our original analytical approach (a negative binomial probability distribution for the dependent variables) is the correct approach. However, as a sensitivity analysis, we have repeated the before and after analysis, excluding all respondents with zero consumption, and log normalizing the consumption data prior to analyses. The findings of the two approaches are very similar.

8. The authors should explain further why it is the case that 'there is no reason to believe that under-reporting would differ by sex.' This is particularly pertinent when using data from a survey that the authors appear to have had no control over the design of.

RESPONSE: We have modified the text here.

Minor comments:
1. The manuscript as a whole should be thoroughly proof-read for errors and awkwardly worded sentences. For example, the first sentence of the "Study design" subsection in Methods needs to be reworked.

**RESPONSE:** We have undertaken thorough proof-reading and editing of the text.

2. What is the reason a weekly average is used? Some justification of this would be useful. it could include a desire to smooth out variation if the within variability is quite high, though reducing variability before analyzing can have issues.

**RESPONSE:** Respondents are asked to complete a diary of previous week’s consumption, and actual day of consumption is not reported in the data set we obtained. We have mentioned this in the methods.

3. Why are the ITS models differenced between country first but the before-and-after analysis not?

**RESPONSE:** These are different analysis strategies. For the ITS, we used differences to control for separate developments. For the before and after analyses, we included actual age and actual deprivation score (rounded to an integer) as dummy-coded variables for each age and each deprivation score in the models, with interaction terms including country by event (introduction of MUP) in the model.

4. For the sub-group analyses, the authors standardize the outcomes to assure comparability of model coefficients. A joint modeling approach where age group is a predictor could be used without standardization. Formal tests can then be constructed.

**RESPONSE:** We consider standardization to be preferable to demonstrate comparisons across the sub-groups, now, age, social grade and deprivation group, as recommended by Beard et al. (Beard, E., Marsden, J., Brown, J., Tombor, I., Stapleton, J., Michie, S., & West, R. (2019). Understanding and using time series analyses in addiction research. *Addiction (Abingdon, England)*, 114(10), 1866–1884. [https://doi.org/10.1111/add.14643](https://doi.org/10.1111/add.14643). This also allows standardized comparability across off and on-license consumption.

5. Some of what appear to be the ITS sub-group results are under the heading "before-and-after analysis" despite not being characterized as such in the statistical analysis section.

**RESPONSE:** We have checked this to ensure that ITS results come in the ITS section, and before and after results in the before and after section.

6. The abbreviation MUP should be defined parenthetically or otherwise in the abstract after minimum unit pricing.

**RESPONSE:** Done

7. Is there evidence of cross-consumption? Have people in Scotland living near the border began consuming in Northern England as a result of the MUP implementation? The analysis as it stands assumes there is no crossover between Scotland and the control.

**RESPONSE:** We cannot analyse this as we do not have the data to do so - but, we have added this issue to the discussion, page 17.

8. The figures on pages 18 and 19 are too small and, consequently, hard to read.

**RESPONSE:** We have adjusted all figures as best as possible to improve readability.

**Reviewer:** 4
Dr. Jim Lewsey, University of Glasgow
Comments to the Author:
I congratulate the authors on a very interesting paper on an important topic. It is well written and is well presented in the main. Controlling the analyses for England is a key strength. However, I have concerns that at present the analyses presented do not provide statistical evidence to support the interpretations you are making about differential MUP effects. Further, I think more reassurance on possible selection/sampling biases not being a major threat to internal validity is required.

RESPONSE: Thank you for all your very helpful comments.

Major:
What is the response rate for KWP? Is it similar between Scotland and England? It is stated that residents from Scotland and 18-34 year olds are oversampled but looking at appendix Figs 1&2 the oversampling looks similar in England and Scotland? Further, it looks like the difference with total population %s is more pronounced for women than men. Given this study is looking at differences between men and women isn’t this a big threat to the internal validity of the study?

RESPONSE: Kantar does not provide data on the proportion of its access panel that take up the offer of completing the survey, and this is not mentioned in other publications using Alcovision data, which we have added to the discussion [e.g.: Stevely AK, Holmes J, Meier PS. Combinations of Drinking Occasion Characteristics Associated with Units of Alcohol Consumed among British Adults: An Event-Level Decision Tree Modeling Study. Alcohol Clin Exp Res. 2021 Mar;45(3):630-637. doi: 10.1111/acer.14560. Epub 2021 Mar 5. PMID: 33666958; Stevely AK, de Vocht F, Neves RB, Holmes J, Meier PS. Evaluating the effects of the Licensing Act 2003 on the characteristics of drinking occasions in England and Wales: a theory of change-guided evaluation of a natural experiment. Addiction. 2021 Sep;116(9):2348-2359. doi: 10.1111/add.15451. Epub 2021 Mar 17. PMID: 33620736.]. The Alcovision data is reported to oversample Scotland compared to England; and, oversample 18-34 year olds in both Scotland and England compared to older age groups, clarified in the methods section. Our analyses are fully weighted: “Weights based on age-sex groups, social class, and geographical region are constructed using UK census data to ensure representativeness of British adults.” Given this, the overweighting of certain age groups only play a role for the CIs for these age groups, but they are no threat for internal validity.

The ITS statistical analyses undertaken do not formally test whether any MUP effect is different between men and women. It is possible to set up the data set as a panel and then use interaction terms in the regression to statistically test for differential effects by sex (and deprivation). I recommend the authors do this (or provide a rationale for why this was not done). Of note, this type of approach seems to have been taken for the ‘before and after analyses’ (models 1&2 include interactions involving MUP&age and MUP&deprivation. Again though, why run separate models for men and women given the aim of this study?

RESPONSE: Using the interaction term, we have tested for the difference between men and women, and demonstrated a difference in both the main analysis and in the new sensitivity analysis, using Northern England as a control.

Results, ITS: claims such as “Associated significant differences in consumption were restricted to women…” are only supported with descriptive interpretations of separate model results for men and women. Interaction tests (see previous point) are needed to support these claims. Same for the paragraph interpreting results shown in Figs 1&2.

RESPONSE: As above.

Results, before and after analyses: some interaction tests were done here but are not reported on. However, separate models were done for men and women so again, at present, the associated interpretation is descriptive and not statistically tested for.

RESPONSE: Now, Figures 4 and 5 are primarily looking at differences by age and deprivation score. We have reported the regression coefficients for the slopes in the figures, and formally tested that the slopes differ between men and women.

18
Minor:
Abstract: Key words; is it reasonable to have 'health inequalities' when the outcome measure in this study is alcohol consumption (not directly health).

RESPONSE: We have amended our key words to incorporate this suggestion.


RESPONSE: Thank you for flagging this 2021 paper which has been added into the discussion.

The labelling and colour scheme in Figs 1&2 of the Appendix make it difficult to know what each line refers to. Can this be made more clear?

RESPONSE: Thank you - yes, we have tried to increase the clarity of these figures.

What was the rationale for grouping age in the categories you chose?

RESPONSE: Given that we had data on young adults, we decided to include an age group of 18–24-year-olds. We used similar age groups in a study of household purchase data (Impact of minimum unit pricing on shifting purchases from higher to lower strength beers in Scotland: Controlled interrupted time series analyses 2015-2020; Drug & Alcohol Review - in Press).

“We adjusted the dependent variables for any seasonality” – how exactly was this done?

RESPONSE: We have added: “We adjusted the dependent variables for any seasonality, using the ratio-to-moving-average method (Ref 23 in the paper)”.

In the ‘before and after analyses’, no explanation is given for why interaction terms are added to models’ 1&2 (and then the results of these are not interpreted).

RESPONSE: We have redone and re-written the before and after analyses, giving the explanation and the use of the interaction terms, page 8.

Typos, etc.:
Abstract: “Although drops in consumption varied by level of residential deprivation, <and differently for men and women>…”; missing word(s)? Sentence does not quite ‘scan’ Appendix: “Deprivatoin” instead of “Deprivation”

RESPONSE: A thorough proof-reading and editing of the revised manuscript has been done.

VERSION 2 – REVIEW

REVIEWER
Mark Robinson
The University of Queensland

REVIEW RETURNED
07-Dec-2021

GENERAL COMMENTS
I’d like to thank the authors for the comprehensive responses to my comments and to those of the other reviewers, and for the substantive changes that they have made to the manuscript. I think the paper is stronger as a result. However, there remains certain aspects of the paper that I think should be considered to enhance clarity and comprehension for the reader. In particular, the framing of the methods needs further consideration. At the moment, the before and after tests are framed as validity checking in the methods.
However, the results stemming from these tests seem to form the basis of the conclusion in the abstract. This, along with the amount of other interpretation afforded to these analyses suggests that it is more than just validation particularly as, in some cases, the results somewhat contradict those from the ITS analyses (e.g. deprivation).

I would strongly encourage a statistician to peer-review the revised version of this paper, particularly the robustness of the consumption percentile analyses, which may be vulnerable due to a small sample in the intervention group.

More specific comments are offered below.

Abstract
The 2nd sentence does not seem to be drawn from the results from the main analyses.
Conclusion section in Word document does not match that in the PDF

Strengths and limitations
4th dot point highlights that representativeness cannot be claimed, yet the methods suggest that the weighting strategy helps to "ensure representativeness of British adults"

Methods
The authors should be commended on the level of methodological detail describing the approach to calculate deprivation scores. However, as the paper is already long Please consider whether such detail is better placed in the supplementary information.

Please consider using a different term to 'data zones' to describe LSOAs in England as this is confusing given this is the official name of the statistical geography in Scotland.

I think it could be made much clearer throughout the paper that the primary outcome measure in the ITS analyses is Scotland minus England. Perhaps a consistent term (e.g. "net effect") might help.

Regression Equation 1. Please consider renaming this to ITS regression equation 1 to avoid any ambiguity / confusion.

Before and after analyses as validity check
As note above, please consider whether this is only validity analyses; the results are discussed as great detail which suggests to me it is being used for the purposes of additional insights. If so, it would be useful to the reader if you state which is the 'primary' analyses and why.

Results
Supplement Table 2
This has a labelling error for Deprivation Group

"Whereas Scottish women" Please add in "in the sample" to avoid misinterpretation

The inclusion of the analyses by consumption level is very useful.

Supplement Table 4
In this table and others, please consider whether stating figures with
3/4 dp is necessary. If not, please reduce the number of dps.

Supplement Table 7
Please check that the zero figures for men are accurate

Interrupted time-series - main findings
(sex of respondent*event, the introduction of MUP ).
This analytical approach is not mentioned in the methods

"6.022 grams per week" - consider if this results is needed to 3dp

"For the 95th percentile for men, the introduction of MUP seemed to be associated with an increase in consumption. For the 95th percentile for women, the confidence intervals crossed zero."
It would be useful if the authors could provide their Perspecton on the robustness of this result. Looking at the sample in certain subgroups suggests that the number of p/pants a the 95th percentile, especially in Scotland, will be rather small. An idea of the uncertainty around the results would be useful.

"confirming the difference in slopes by age between men and women."
Should age be deprivation in this sentence?

Figure 3
The label should more clearly state the units of the standardised coefficients.
I agree with reviewer comment on the original version that Figure 3 is almost unreadable because of its size

"it seemed that there were greater associated drops …" The phrase 'seemed that' is used frequently in the revised version of the manuscript. It would be more convincing for the reader if the best estimate of the effect was provided with an idea of uncertainty. I don’t see much value of including this phrase so it could just be removed.

Before and after analyses as validity check
"We undertook the before-and-after analyses" The first sentence is explained in the methods and so does not need repeating here.

"The coefficient for the interaction term, sex=men by age" Is this an interaction term. If so, should an asterisk be used for consistency?

Discussion
Specify Scotland in the first sentence.

Again, the results from the validation analyses is headlines in the first para of the discussion. Please consider if this is appropriate, particularly in terms of the robustness/uncertainty around the ‘heaviest 5%’ finding.

"Based on the interrupted time series analyses…" How might this difference between the two methods be explained?

"It has been suggested that some very heavy drinkers……" Please consider the recent report by MESAS for relevance:
"Unlike the household..."
Please consider the following MESAS report for relevance on cross-border shopping:
http://www.healthscotland.scot/publications/minimum-unit-pricing-evaluation-compliance-study

"...in emergency room visits" Please specify the outcome measure - was it alcohol-related visits?

STROBE checklist
This states that no sensitivity analyses was performed.

REVIEWER
Mark Meyer
Georgetown University, Mathematics and Statistics

REVIEW RETURNED
10-Dec-2021

GENERAL COMMENTS
Statistical Review for BMJ Open manuscript bmjopen-2021-054161 R1

Differential impact of minimum unit pricing on alcohol consumption between Scottish men and women: controlled interrupted time-series analysis."

I would rst like to thank the authors for their work on this revision. They have clearly put in a lot of e ort and the manuscript is much improved. There is, however, one outstanding major issue held over from the original submission. To summarize, the "grams of alcohol consumed" variable, even after rounding, represents realizations from a continuous random variable and must be analyzed accordingly. The log-normal analysis the authors currently perform as their sensitivity analysis should be the primary analysis for this variable. The results from this model should be expanded and included in the manuscript including estimates and con dence intervals along with the existing graphical depictions.

The negative binomial analysis should be removed from the manuscript and the accompanying material entirely. I will now discuss in more detail the reason why the negative binomial analysis is not appropriate.

In my rst review, I noted that the use of the negative binomial model for a continuous random variable is incorrect. The authors’ response con rmed that the outcome variable was continuously sampled. However, they go on to state that 'we have treated consumption (rounded to an integer) as a count variable (counting grams of alcohol consumed)." This analysis is still incorrect. Rounding a continuously measured variable to an integer value does not make it a realization from a categorical or discrete random variable. It still arises from a continuous random variable. The integers are a subset of the reals, so the support of the random variable has not changed. But more to the point, counting processes are not de ned in this fashion. The rounded grams variable is no more counting grams than a rounded temperature variable counts degrees or a rounded height measurement counts inches. Rounding is a common convention for these types of measurements, particularly self-reported ones, but that does not transform them into realizations from discrete random variables.

The authors’ use of the negative binomial model is, presumably, in lieu of the poisson model (as they authors’ noted in their original
submission as well as the revision, the variable is `highly dispersed" suggesting they would have otherwise used Poisson regression.

Poisson processes are carefully de ned random variables that should satisfy the Poisson Postulates (Casella and Berger, 2002, 135{136). There are several features to Poisson processes but a key assumption is that the process is counting a series of independent events occurring in disjoint intervals. Or, as Pagano and Gauvreau put it better, "[t]he events occur independently both within the same interval and between consecutive intervals" (Pagano and Gauvreau, 2000, 173). The dependent variable `grams of alcohol consumed’ does not arise from a counting of independent events occurring over time or space. There may be a way to de ne events in the context of consuming alcohol that would justify analysis with Poisson or negative binomial models. But the variable as de ned does not satisfy this key condition. For a technical reviews of the Poisson postulates, please refer to Casella and Berger (2002), pages 135{136 or Wackerly, Mendenhall, and Scheaffer (2008), pages 131{135. For a less technical review, please see Pagano and Gauvreau (2000), pages 172{175.

Using statistical models for discrete random variables on continuously measured random variables is problematic because of the implied variances. The authors were correct to note the variable exhibits overdispersion when viewed from the lens of the Poisson model realizations from a continuous random variable are very likely to exhibit such dispersion if incorrectly viewed as count processes. The Poisson model assumption that the variance equals the mean is obviously much too strong an assumption for this data. But the negative binomial model also induces a strong assumption about the relationship between the mean and variance: the variance is a quadratic function of the mean. Speci cally, the induced variance from negative binomial regression is

\[ \text{Var}(Y) = \mu + 2\theta; \]

where \(\theta\) is the dispersion parameter, for more detail on the derivation of and assumptions for negative binomial regression, please see Agresti (2013), pages 552{555. This is still a strong assumption particularly for data that does not actually arise from a Poisson process exhibiting extra-Poisson variability. The variability in the authors’ negative binomial model is therefore incorrectly estimated and inferences resulting from this analysis cannot be trusted.

The log-normal sensitivity analysis is a more appropriate treatment of the data and I thank the authors for performing it for this revision. However, the results that they present are limited and only include graphical

results. No numerical results are given, so no formal comparison can be made regarding inference. The scales on the graphs are di erent enough between the analyses that it is di cult to judge how the log-transformed analysis impacts variance estimation. Further, I would contend that the results do di er in some meaningful ways. In particular, the on-trade curve is at and negative in the sensitivity
analysis. This is not in the negative binomial analysis. As I noted above, the log-normal analysis should replace the negative binomial analysis as the primary analysis for this outcome. The details of this model need to be expanded in section titled 'Before and after analyses as validity check.' The results section and discussion should be updated accordingly to reflect the log-normal analysis. Finally, the negative binomial analysis should be removed entirely, from both the manuscript and the supporting material. If the authors wish to perform a sensitivity analysis to the log-normal analysis, they could try a root-normal model (taking the square root, instead of the log, to normalize) or perform a gamma regression on the data using the original scale.

References


VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Mark Robinson, The University of Queensland Comments to the Author:

I'd like to thank the authors for the comprehensive responses to my comments and to those of the other reviewers, and for the substantive changes that they have made to the manuscript. I think the paper is stronger as a result. However, there remains certain aspects of the paper that I think should be considered to enhance clarity and comprehension for the reader. In particular, the framing of the methods needs further consideration. At the moment, the before and after tests are framed as validity checking in the methods. However, the results stemming from these tests seem to form the basis of the conclusion in the abstract. This, along with the amount of other interpretation afforded to these analyses suggests that it is more than just validation particularly as, in some cases, the results somewhat contradict those from the ITS analyses (e.g. deprivation).

RESPONSE: Thank you. We have re-framed the before and after analyses as secondary analyses to investigate in more detail the potential impact of MUP by individual age of respondent and by individual residential deprivation ranking of where the respondent lived. We have identified the ITS as primary analysis.
I would strongly encourage a statistician to peer-review the revised version of this paper, particularly the robustness of the consumption percentile analyses, which may be vulnerable due to a small sample in the intervention group.

RESPONSE: Thank you. Reviewer 3 is a statistician and has not raised the robustness of the interrupted time series consumption percentile analyses as a concern. The analysis considers differences between Scotland and England for each of the 19 percentiles for each of the two time periods prior to MUP and from MUP onwards, separately for men and women. In each percentile, this means before MUP the average differences between 633 Scottish residents and 4046 English residents and after MUP, the differences between 121 Scottish residents and 805 English residents, all split roughly equally between men and women. The wide 95% confidence intervals reflect the relatively small sample sizes, nevertheless, allowing us to interpret the findings. We have added a summary of the number of respondents per percentile as a footnote to a new Supplement Table 9, Page 22, and added a note about this in the discussion.

[Reviewer 3 kindly addressed the issue of the before and after analyses, which we have corrected as proposed by the reviewer].

More specific comments are offered below.

Abstract

The 2nd sentence does not seem to be drawn from the results from the main analyses.

Conclusion section in Word document does not match that in the PDF

RESPONSE: The second sentence, ‘Associated reductions were larger for women than for men and were greater amongst heavier as opposed to lighter drinkers, except for the 5% of heaviest drinking men, who seemed to increase their consumption’ is derived from the primary ITS analysis. We have checked and corrected any discrepancies between the word and pdf documents.

Strengths and limitations

4th dot point highlights that representativeness cannot be claimed, yet the methods suggest that the weighting strategy helps to “ensure representativeness of British adults”

RESPONSE: In the methods section, we deleted the phrase “to ensure representativeness of British adults”.

Methods
The authors should be commended on the level of methodological detail describing the approach to calculate deprivation scores. However, as the paper is already long Please consider whether such detail is better placed in the supplementary information.

RESPONSE: Where possible, we have moved text to the supplement, indicating this in the text of the main paper.

Please consider using a different term to 'data zones' to describe LSOAs in England as this is confusing given this is the official name of the statistical geography in Scotland.

RESPONSE: We have changed to 'data areas' (now in supplement).

I think it could be made much clearer throughout the paper that the primary outcome measure in the ITS analyses is Scotland minus England. Perhaps a consistent term (e.g. 'net effect') might help.

RESPONSE: Thank you. We have added, as appropriate, 'net effect' throughout.

Regression Equation 1. Please consider renaming this to ITS regression equation 1 to avoid any ambiguity / confusion.

RESPONSE: We have placed the equations in the supplement, together with the syntax and renamed accordingly.

Before and after analyses as validity check As note above, please consider whether this is only validity analyses; the results are discussed as great detail which suggests to me it is being used for the purposes of additional insights. If so, it would be useful to the reader if you state which is the 'primary' analyses and why.

RESPONSE: We have renamed the ITS as the primary analysis and the before and after as the secondary analysis.

Results
Supplement Table 2
This has a labelling error for Deprivation Group
RESPONSE: Corrected.

"Whereas Scottish women" Please add in "in the sample" to avoid misinterpretation
RESPONSE: Done.

The inclusion of the analyses by consumption level is very useful.
RESPONSE: Thank you.

Supplement Table 4
In this table and others, please consider whether stating figures with 3/4 dp is necessary. If not, please reduce the number of dps.
RESPONSE: In the text, outside of the tables, we have reduced numbers of dps throughout, as appropriate. To be consistent throughout the tables, we have left the tables with 3 dps, since one of the coefficients, for time (weeks) in some cases requires 3 dps.

Supplement Table 7
Please check that the zero figures for men are accurate
RESPONSE: In Supplement Tables 7 and 8, have added that the coefficients of ‘0.000’ are the reference group.

Interrupted time-series - main findings
(sex of respondent*event, the introduction of MUP).
This analytical approach is not mentioned in the methods
RESPONSE: We have now included.

"6.022 grams per week" - consider if this results is needed to 3dp
RESPONSE: As above for dps.

"For the 95th percentile for men, the introduction of MUP seemed to be associated with an increase in consumption. For the 95th percentile for women, the confidence intervals crossed zero."

It would be useful if the authors could provide their Perspective on the robustness of this result. Looking at the sample in certain subgroups suggests that the number of p/pants a the 95th percentile, especially in Scotland, will be rather small. An idea of the uncertainty around the results would be useful.
RESPONSE: Please see response to your second comment above. We have added a summary of the number of respondents per percentile as a footnote to a new Supplement Table 9, Page 22, and added a note about this in the discussion.
"confirming the difference in slopes by age between men and women."

Should age be deprivation in this sentence?

RESPONSE: We have corrected the typing error.

Figure 3

The label should more clearly state the units of the standardised coefficients.

I agree with reviewer comment on the original version that Figure 3 is almost unreadable because of its size

RESPONSE: We have clarified the label and have redone the figure to make it more readable. [Part of the problem is the way that it has been reproduced on the page in the automated submission process].

"It seemed that there were greater associated drops …" The phrase 'seemed that' is used frequently in the revised version of the manuscript. It would be more convincing for the reader if the best estimate of the effect was provided with an idea of uncertainty. I don't see much value of including this phrase so it could just be removed.

RESPONSE: We have edited the text and removed 'seemed that' throughout.

Before and after analyses as validity check "We undertook the before-and-after analyses" The first sentence is explained in the methods and so does not need repeating here.

RESPONSE: We have deleted this in the results.

"The coefficient for the interaction term, sex=men by age" Is this an interaction term. If so, should an asterisk be used for consistency?

RESPONSE: Yes, corrected.

Discussion

Specify Scotland in the first sentence.

RESPONSE: Done.

Again, the results from the validation analyses is headlines in the first para of the discussion. Please consider if this is appropriate, particularly in terms of the robustness/uncertainty around the 'heaviest 5%' finding.

RESPONSE: The heaviest drinking analysis is based on the ITS. See previous response to comments about the uncertainty.
"Based on the interrupted time series analyses..." How might this difference between the two methods be explained?

RESPONSE: Following the re-doing on the secondary before and after (B&A) analysis, the findings between the primary ITS and the secondary B&A are more consistent.

"It has been suggested that some very heavy drinkers......" Please consider the recent report by MESAS for relevance: https://www.publichealthscotland.scot/media/8201/the-impact-of-mup-among-people-who-are-alcohol-dependent-and-accessing-treatment-services-briefing.pdf

RESPONSE: Thank you. We have included.

"Unlike the household...."

Please consider the following MESAS report for relevance on cross-border shopping:
http://www.healthscotland.scot/publications/minimum-unit-pricing-evaluation-compliance-study

RESPONSE: Thank you. We have included.

"...in emergency room visits" Please specify the outcome measure - was it alcohol-related visits?

RESPONSE: Changed to ‘Alcohol-related emergency department visits’.

STROBE checklist

This states that no sensitivity analyses was performed.

RESPONSE: Adjusted STROBE checklist.

REVIEWER 3

I would first like to thank the authors for their work on this revision. They have clearly put in a lot of effort and the manuscript is much improved. There is, however, one outstanding major issue held over from the original submission. To summarize, the ‘grams of alcohol consumed’ variable, even after rounding, represents realizations from a continuous random variable and must be analyzed accordingly. The log-normal analysis the authors currently perform as their sensitivity analysis should be the primary analysis for this variable. The results from this model should be expanded and included in the manuscript including estimates and confidence intervals along with the existing graphical depictions. The negative binomial analysis should be removed from the manuscript and the accompanying material entirely. I will now discuss in more detail the reason why the negative binomial analysis is not appropriate. In my first review, I noted that the use of the negative binomial model for a
continuous random variable is incorrect. The authors' response confirmed that the outcome variable was continuously sampled. However, they go on to state that "we have treated consumption (rounded to an integer) as a count variable (counting grams of alcohol consumed)." This analysis is still incorrect. Rounding a continuously measured variable to an integer value does not make it a realization from a categorical or discrete random variable. It still arises from a continuous random variable. The integers are a subset of the reals, so the support of the random variable has not changed. But more to the point, counting processes are not defined in this fashion. The rounded grams variable is no more counting grams than a rounded temperature variable counts degrees or a rounded height measurement counts inches. Rounding is a common convention for these types of measurements, particularly self-reported ones, but that does not transform them into realizations from discrete random variables. The authors' use of the negative binomial model is, presumably, in lieu of the poisson model (as they authors' noted in their original submission as well as the revision, the variable is "highly dispersed" suggesting they would have otherwise used Poisson regression).

Poisson processes are carefully defined random variables that should satisfy the Poisson Postulates (Casella and Berger, 2002, 135-136). There are several features to Poisson processes but a key assumption is that the process is counting a series of independent events occurring in disjoint intervals. Or, as Pagano and Gauvreau put it better, "the events occur independently both within the same interval and between consecutive intervals" (Pagano and Gauvreau, 2000, 173). The dependent variable "grams of alcohol consumed" does not arise from a counting of independent events occurring over time or space. There may be a way to define events in the context of consuming alcohol that would justify analysis with Poisson or negative binomial models. But the variable as defined does not satisfy this key condition. For a technical review of the Poisson postulates, please refer to Casella and Berger (2002), pages 135-136 or Wackerly, Mendenhall, and Scheaffer (2008), pages 131-135. For a less technical review, please see Pagano and Gauvreau (2000), pages 172-175. Using statistical models for discrete random variables on continuously measured random variables is problematic because of the implied variances. The authors were correct to note the variable exhibits overdispersion when viewed from the lens of the Poisson model|realizations from a continuous random variable are very likely to exhibit such dispersion if incorrectly viewed as count processes. The Poisson model assumption that the variance equals the mean is obviously much too strong an assumption for this data. But the negative binomial model also induces a strong assumption about the relationship between the mean and variance: the variance is a quadratic function of the mean. Specifically, the induced variance from negative binomial regression is \[ \text{Var}(Y) = \mu + \phi \mu^2 \] where \( \mu = \text{E}(Y) \) and \( \phi \) is the dispersion parameter, for more detail on the derivation of and assumptions for negative binomial regression, please see Agresti (2013), pages 552-555. This is still a strong assumption particularly for data that does not actually arise from a Poisson process exhibiting extra-Poisson variability. The variability in the authors' negative binomial model is therefore incorrectly estimated and inferences resulting from this analysis cannot be trusted.

The log-normal sensitivity analysis is a more appropriate treatment of the data and I thank the authors for performing it for this revision. However, the results that they present are limited and only include graphical 1 results. No numerical results are given, so no formal comparison can be made regarding inference. The scales on the graphs are different enough between the analyses that it is difficult to judge how the log-transformed analysis impacts variance estimation. Further, I would contend that the results do differ in some meaningful ways. In particular, the on-trade curve is at and negative in the sensitivity analysis. This is not in the negative binomial analysis. As I noted above, the log-normal analysis should replace the negative binomial analysis as the primary analysis for this outcome. The details of this model need to be expanded in section titled 'Before and after analyses as validity check." The results section and discussion should be updated accordingly to reflect the log-normal analysis. Finally, the negative binomial analysis should be removed entirely, from both the manuscript and the supporting material. If the authors wish to perform a sensitivity analysis to the log-normal analysis, they could try a root-normal model (taking the square root, instead of the log, to normalize) or perform a gamma regression on the data using the original scale.
RESPONSE: Thank you. We are very grateful for this explanation. We have removed all material related to the negative binomial analysis, and replaced it using the log-normal analysis, adjusting the methods, results and discussion accordingly. As sensitive analysis, we used a root-normal model.

VERSION 3 – REVIEW

REVIEWER | Mark Robinson  
The University of Queensland
--- | ---
REVIEW RETURNED | 04-Mar-2022

GENERAL COMMENTS | I would like to thank the authors for their major revision to their paper. The reframing of the primary and secondary analyses has helped provide greater clarity in the results and associated interpretations. One final minor revision I would request is that the observed increase in consumption associated with MUP among the heaviest 5% of male drinkers is attempted to be explained. The authors report it as a key finding in the abstract and conclusion, but there is currently no attempt to explain how reduced affordability of alcohol might lead to an “increase” in consumption among those spending the most amount of money on alcohol (and typically with the least disposable income).

Kind regards,

Mark Robinson

REVIEWER | Mark Meyer  
Georgetown University, Mathematics and Statistics
--- | ---
REVIEW RETURNED | 12-Mar-2022

GENERAL COMMENTS | The authors have greatly improved the manuscript and I thank them for their diligence in responding to comments. My only remaining comment is that the manuscript should be thoroughly proofread for spelling errors. For example, I noticed Scotland was misspelled in a few places.

RESPONSE: Many thanks. We have done our best to respond to this in the discussion by adding:

“We can only speculate about the reasons for the increase in the five per cent of the heaviest drinking men. Several studies have found that overall, heavier drinkers, including people with alcohol use disorders, react less to price than the general population, i.e., they react more price inelastic and their consumption is determined by other factors (see reviews and meta-analyses46, 47). However, while this may explain lower reductions, it cannot explain an increase in consumption. Such a polarization with increasing consumption of the heaviest drinkers in overall decreasing consumption levels has now been observed in several studies, often in adolescents and young adults48, 49. These studies indicate that such polarization means a deviation from the standard collective theory of all subgroups changing in the same direction,50 but fall short on good explanations as to why this is the case.”
Reviewer: 3
Dr. Mark Meyer, Georgetown University
Comments to the Author:
The authors have greatly improved the manuscript and I thank them for their diligence in responding to comments. My only remaining comment is that the manuscript should be thoroughly proofread for spelling errors. For example, I noticed Scotland was misspelled in a few places.
RESPONSE: Thank you. We have checked and corrected all typing mistakes.