PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Road traffic casualties in Great Britain at daylight savings time transitions: a causal regression discontinuity design analysis
AUTHORS	Singh, Ramandeep; Sood, Rohan; Graham, Daniel

VERSION 1 – REVIEW

REVIEWER	Martin, JL Université de Lyon, IFSTTAR/TS2/Umrestte
REVIEW RETURNED	20-Jul-2021

GENERAL COMMENTS	At a time when several European states are preparing to change their legislation on summer time, the subject of this article is certainly very relevant. The article appears to be very well written, the objectives are clear, the method is well explained, and the data used are correctly described. In particular, the statistical method used (RDD) appears particularly appropriate, and the conditions for its use have been well controlled. The discussion of the different hypotheses concerning the possible sleep disturbance or the effect of the change of brightness appears coherent in view of the results obtained. I have only two very minor comments: - I suggest that the line relating to autumn should appear in Table 2, only that relating to spring appears, even if the % change is not significant - Page 4, line 18, Ordinance should be Ordnance I therefore unreservedly support the publication of this article.

REVIEWER	Uttley, Jim The University of Sheffield, School of Architecture
REVIEW RETURNED	03-Aug-2021

GENERAL COMMENTS	The paper reports an analysis of road traffic casualties in Great Britain, with the aim of identifying whether there is a causal relationship between daylight savings transitions and the number of road casualties. The topic is important and the approach taken by the authors seems generally appropriate. I have a number of major and minor comments and points of clarification however. Major comments: The 3 week period that is used either side of the clock change feels like a long period when looking for effects of daylight saving transition. For example I would expect the sleep hypothesis to apply in the first few days before / after a transition, but 3 weeks

	away from the clock change, I would expect people to have adjusted their sleep patterns. Use of the label 'Twilight' to cover the 00:00-06:59 period is a bit misleading. Twilight covers only a short part of the day, and much of the time period used here will be in darkness not twilight. With the number of models used and number of significance tests carried out, is there a danger of obtaining a number of Type I errors, particularly as an alpha of 0.1 is being used rather than the conventional 0.05? Should you correct for multiple testing? Tables 3 and 4 suggest only a small number of models with fatalities as the outcome measure passed the specification tests. Compared with the number of all casualty models that passed these tests, this suggests the smaller number of fatalities impacts on whether the model passes this screening. The small number of fatalities therefore implies any conclusions that can be drawn are limited. Absence of evidence of an effect is not the same as evidence of absence. What is your view on this? An aggregate model for each Northing and Easting band (i.e. collapsing across time periods) would be useful so that the overall effect of the band can be more easily discerned. Also, can you explain the bands - do lower bands represent more Northern and Eastern locations, or the other way round? You test the sleep hypothesis by assuming people will be sleepier in the Morning peak period, and therefore look for an increase in road casualties in that period. However, how do you know any sleepiness will manifest itself immediately in the morning, and not later in the day a person grows more tired due to an earlier than normal waking time? Can you say something about how much extra time would be spent in darkness for western or northern locations? Are we talking a few minutes or a few hours? When I compare the onset of civil twilight in Plymouth and London, for example, there is only a 20 minute difference, suggesting differences in amount of darkness experienced by locations in East vs West may not be very significant. Is it possible that the sleep and light hypotheses may be countering each other? In the Spring transition, people may feel sleepier in the mornings after the clock change, but they are more likely to be driving in daylight rather than darkness. Similarly, in Autumn, people will get an extra hour in bed so may be less sleepy after the clock change, but are more likely to be travelling in darkness than daylight in the morning. You test the light hypothesis by looking for changes in road casualties in the PM peak and night time periods. However, I'm not sure we would expect any effect in the night time period as this period is in darkness both before and after the clock change - there is little to no difference in ambient light before and after. To assess both light and sleep hypotheses it seems to me that you would need to compare changes in casualties in the relevant period (e.g. PM peak for light, AM peak for sleep) with changes in other periods, to account for any changes in casualty numbers due to other factors unrelated to these two hypotheses (for example, school vacation periods could consistently fall on one side of the clock change, potentially influencing casualty rates due to increases (or decreases) in road traffic. Equally, weather conditions could be another conceivable factor, with potentially worse weather conditions falling before the Spring clock change and after the Autumn clock change, which could offset changes in road casualties due to the light and sleep hypotheses).

	Minor comments: Check spelling of “casual” throughout the paper Line 24-26, page 4 - there are a few recent studies that would support a causal link between changes in light conditions and road collisions, e.g. Uttley, J., & Fotios, S. (2017). The effect of ambient light condition on road traffic collisions involving pedestrians on pedestrian crossings. Accident Analysis & Prevention, 108, 189-200. Fotios, S., Robbins, C. J., & Uttley, J. (2021). A comparison of approaches for investigating the impact of ambient light on road traffic collisions. Lighting Research & Technology, 53(3), 249-261. Raynham, P., Unwin, J., Khazova, M., & Tolia, S. (2020). The role of lighting in road traffic collisions. Lighting Research & Technology, 52(4), 485-494. “Collisions” or “Crashes” is preferred to “Accidents”, e.g. see Davis, R. M., & Pless, B. (2001). BMJ bans “accidents”: Accidents are not unpredictable. Can the authors clarify what is meant by ‘local bandwidth’, in the context of the regression discontinuity method, particularly the “optimised local bandwidth” as used in line 47, page 4? In particular, why different bandwidths are used for different models? This could raise questions about consistency and ‘cherry picking’ of results to fit with requirements. Line 48, page 4 - Can you say more about what was missing from these 0.02% of observations? Can you explain why Bank Holiday records were removed? Can the code and data for the analysis be made openly accessible?

REVIEWER	Hanley, JA McGill University, Department of Epidemiology and Biostatistics
REVIEW RETURNED	27-Oct-2021

GENERAL COMMENTS	The normal error model has advantages over Poisson etc, in that the variances will be more accurately measured (and not model-based) But you do have small counts with the time sliced so finely any way to have a falsification test? maybe slide time by 2 weeks and pretend DST kicked in earlier later show middle of the day results separately, and show all data graphically -- raw as well as fitted models. and by period of day... readers need to understand why in Spring, raw counts are higher post than pre, but model fit suggests otherwise.. seeing the data is more likely to help convince readers that waht you did/found is reasonable. Disc. of USA results (where FARS system is more complete) -- and any other countries also, would help.

	Giving more of the coefficients of the fitted models would also help us judge if fits seem sensible and accord with what would be expected. any place for a finely matched analysis (eg. Mantel-Haenszel, with difference in differences)? What happens if do this for July or September or May? (gives some sense of variations under the null) Are you over-selling 'causal'. What makes yours more causal than others? It is still regression, and using regression models to turn back the clock ('setting' things this week to what they 'would have been') is easier said that done. The split b/w light and sleep is a plus.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 -	***Dr. JL Martin, Université de Lyon***	
	At a time when several European states are preparing to change their legislation on summer time, the subject of this article is certainly very relevant. The article appears to be very well written, the objectives are clear, the method is well explained, and the data used are correctly described. In particular, the statistical method used (RDD) appears particularly appropriate, and the conditions for its use have been well controlled. The discussion of the different hypotheses concerning the possible sleep disturbance or the effect of the change of brightness appears coherent in view of the results obtained. I have only two very minor comments:	-
1	I suggest that the line relating to autumn should appear in Table 2, only that relating to spring appears, even if the % change is not significant	In Table 2, we have now included rows to make clear that the RDD results for the Autumn fatalities, Spring total casualties and Spring fatalities are not significant.
2	Page 4, line 18, Ordinance should be Ordnance	This correction has now been made on page 4.
	I therefore unreservedly support the publication of this article.	We thank the reviewer for their endorsement of our work.
Reviewer 2 -	***Dr. Jim Uttley, The University of Sheffield***	
	The paper reports an analysis of road traffic casualties in Great Britain, with the aim of identifying whether there is a causal relationship between daylight savings transitions and the number of road casualties. The topic is important and the approach taken by the authors seems generally	-

	appropriate. I have a number of major and minor comments and points of clarification however.	
	Major comments:	-

Number	Reviewer comment	Author response
1	The 3 week period that is used either side of the clock change feels like a long period when looking for effects of daylight saving transition. For example I would expect the sleep hypothesis to apply in the first few days before / after a transition, but 3 weeks away from the clock change, I would expect people to have adjusted their sleep patterns.	The 3 week period quoted in Section 2.2 is the original data set window to ensure that we have enough data points for the optimal bandwidth selection process. To clarify this in Section 2.2, we have mentioned that "Three weeks is chosen to provide enough data for the optimised local bandwidth to be calculated during the RDD modelling". In RDD modelling, the concept of optimal bandwidth selection based on minimising the mean squared error of the average treatment effect is well-established in the RDD literature (Cattanaeo and Vazquez-Bare, 2016; Calanico et al., 2015; Imbens and Kalyanaraman, 2012). The optimal bandwidth associated with each regression model with a significant average treatment effect is given in Tables 2, 3, and 4. As shown in the tables, the maximum bandwidth is 78 time periods, which corresponds to 7.8 days either side of the cutoff, while the minimum bandwidth is 18 time periods or 1.8 days either side of the cutoff, and the mean bandwidth is approximately 44 time periods or 4.4 days either side of the cutoff. The bandwidth range is therefore reasonably short such that we can observe potential impacts of sleep and light changes in the immediate vicinity of the transition. For further information, we have also added the total number of casualties before and after the cutoff for every model in Tables 2, 3, and 4.
2	Use of the label ‘Twilight’ to cover the 00:00-06:59 period is a bit misleading. Twilight covers only a short part of the day, and much of the time period used here will be in darkness not twilight.	We thank the reviewer for picking this up, and we have now replaced "Twilight" with the term "Overnight".
3	With the number of models used and number of significance tests carried out, is there a danger of obtaining a number of Type I errors, particularly as an alpha of 0.1 is being used rather than the	In the engineering literature, a significance cutoff of alpha=0.1 is often adopted, however, we are happy to tighten the threshold to alpha=0.05. The higher threshold for statistical significance does not

	conventional 0.05? Should you correct for multiple testing?	appreciably alter the overall general findings of the paper, as the impact of daylight savings transitions on road casualties remains minimal. At both transitions combined, there are now approximately 2.9-3.5 fewer total casualties (compared to the previous range of 3.3-3.9) and 0.10-0.24 fewer fatalities (compared to the previous range of 0.25-0.36), when considering the significant models only (note: further on in this response, we detail that we now no longer report the summary values in this way). We have updated Tables 2, 3, and 4 with the new 95% significance level. We are grateful to the reviewer for making the very good point that multiple testing could be an issue. Though we do not believe that corrections should be made to the p-values in each model, we do believe that uncertainty should be quantified when reporting the summary findings of the analysis. We discuss our rationale as follows. Correcting for multiple testing is a concept that is applied in cases when multiple outcome variables are compared and interpreted in a joint manner (i.e. as a family of tests). Corrections typically involve making adjustments to reduce the critical p-value threshold for each model estimated. In our analysis, we define and interpret the models relating total casualties and fatalities in an independent manner, and the models are generated under the assumption that each band and time-period sample is independent. To ensure that this assumption is met, we have addressed potential confounding by including time-varying covariates, and adjusting for heteroskedasticity and auto-correlation within the model framework; these adjustments result in higher standard errors and lower statistical significance of the average treatment effect. Furthermore, we have undertaken a number of specification tests (for polynomial order, bandwidth, and placebo testing). As a result of our stringent model specifications that address potential bias in each model, we do not believe that additional blanket adjustments to the p-value for each model is appropriate, however, we do believe that a correction should be applied when we report on the summary of our findings.

Number	Reviewer comment	Author response
		We report on the combined impact of casualties (total casualties and fatalities separately) across both the Spring and Autumn transitions. In the previous version of the paper, we summed the average treatment effects from significant models over all time periods across the Spring and Autumn transitions, separately for the northings and easting bands. Based on the reviewer's comment, we now make an adjustment in the reporting of the summary values to account for uncertainty that could arise from pooling all the results together. In the new revision of the paper, we generate 95% bootstrap confidence intervals using 10,000 iterations as per the bias corrected and accelerated bootstrap method (DiCiccio and Efron, 1996; Efron, 1987). The statistic of interest that we bootstrap is calculated in two steps: (1) We sum all average treatment effects in the regional time of day models over the Spring and Autumn transitions combined. Two estimates are generated: one for Easting band segmentation and one for Northing band segmentation. (2) We calculate the mean of the Easting and Northing band values, and this is taken as the estimated combined number of casualties over the Spring and Autumn transitions. We perform this procedure for fatalities and total casualties separately. This is documented in Section 4.4.
4	Tables 3 and 4 suggest only a small number of models with fatalities as the outcome measure passed the specification tests. Compared with the number of all casualty models that passed these tests, this suggests the smaller number of fatalities impacts on whether the model passes this screening. The small number of fatalities therefore implies any conclusions that can be drawn are limited. Absence of evidence of an effect is not the same as evidence of absence. What is your view on this?	We agree with the reviewer that the reason why fewer fatality models have a significant average treatment effect could be that there are fewer fatalities either side of the cutoff. We believe that it is important to mention that the majority of models do not indicate a significant change in fatalities at the cutoff (i.e. absence of evidence of an effect), but we do distinguish that this does not indicate evidence of absence. We have now mentioned these points in Section 4.4.
5	An aggregate model for each Northing and Easting band (i.e. collapsing across time periods)	We have included Figure 1 in Section 3 to illustrate the band locations, and we have also added the

	would be useful so that the overall effect of the band can be more easily discerned. Also, can you explain the bands - do lower bands represent more Northern and Eastern locations, or the other way round?	following text on page 6: "As shown in the figure, higher band numbers represent more northern and more eastern locations." In terms of reporting on models at aggregate easting and northing bands over all time periods, we initially generated these models, however, we found that the interpretation of the models was too convoluted. There are distinct effects in the morning at sunrise times and in the evening at sunset times, and the effects would not be able to be disentangled. We therefore made the decision to report on the models split by time period only as these have fewer conflicting and/or compounded effects.
6	You test the sleep hypothesis by assuming people will be sleepier in the Morning peak period, and therefore look for an increase in road casualties in that period. However, how do you know any sleepiness will manifest itself immediately in the morning, and not later in the day a person grows more tired due to an earlier than normal waking time?	We thank the reviewer for raising this point. We have now amended the discussion in Section 4 to include consideration for the sleep hypothesis to manifest throughout the day at both transitions.

Number	Reviewer comment	Author response
7	Can you say something about how much extra time would be spent in darkness for western or northern locations? Are we talking a few minutes or a few hours? When I compare the onset of civil twilight in Plymouth and London, for example, there is only a 20 minute difference, suggesting differences in amount of darkness experienced by locations in East vs West may not be very significant.	The difference in civil twilight sunset/sunrise times between Lowestoft the most easterly location in Great Britain and Aberystwyth the associated most westerly location in the same latitude band is approximately 23 minutes. While this may not seem substantial, the extra darkness at the Autumn transition occurs in the afternoon peak times for traffic, with civil twilight sunset time at 17:00 and 17:24 in Lowestoft and Aberystwyth, respectively. Therefore the difference in light during the peak hour could have a sizeable impact in the Autumn PM period. In terms of north vs south differences, at the Spring transition, locations in the north have longer periods of sunlight (earlier sunrise and later sunset) than locations in the south, for locations in the same longitude band. At the Autumn transition, locations in the north have shorter periods of sunlight

			(later sunrise and earlier sunset) than locations in the south. We have detailed the full implications of the light changes in terms of north/south and east/west differences for each transition in our response to comment 8 below.
	8	Is it possible that the sleep and light hypotheses may be countering each other? In the Spring transition, people may feel sleepier in the mornings after the clock change, but they are more likely to be driving in daylight rather than darkness. Similarly, in Autumn, people will get an extra hour in bed so may be less sleepy after the clock change, but are more likely to be travelling in darkness than daylight in the morning.	We have now improved our descriptions of the changes in sleep and light including discussions of all potential conflicts as follows (note all sunrise/sunset times quoted refer to civil twilight times): At the Spring transition, clocks go forward 1 hour, resulting in an hour less sleep - In the morning, civil twilight sunrise times change from approximately 5-5:30am to 6-6:30am across Britain. There is an hour less sleep and mornings are darker by an hour before 6-6:30am. These conditions could result in a compounding effect of the sleep and light hypotheses, likely resulting in an increase in casualties. In terms of regional effects, in the most western locations of Great Britain, the sun rises approximately 23 minutes later than the most eastern locations. As the civil twilight sunrise times coincide with the beginning of the morning peak in traffic, there could be a possibility of more casualties in the darker western locations compared to the east. Furthermore, sunrise at the most southern locations occurs approximately 20 minutes after the most northern locations, and so there could be a possibility of more casualties in the south relative to the north. - Civil daylight occurs throughout the AM peak (7-10am), inter-peak (10am-4pm), and PM peak (4pm-7pm), and so the light hypothesis is not applicable in these time periods. The sleep hypothesis is applicable, and sleepiness could manifest throughout the day, leading to potential increases in casualties - In the evening, civil twilight sunset times change from approximately 7-7:30pm to 8-8:30pm across Great Britain. There is an hour less sleep throughout the day, but evenings are lighter by an hour in the off-peak travel time after 7-7.30pm, therefore resulting in a potential conflict of the sleep and light

		hypotheses. In terms of regional effects of the light hypothesis, the most western locations experience sunset approximately 23 minutes after the most eastern locations, so there may be potential for increased casualties in the east compared to the west due to the light hypothesis. There is minimal difference (approximately 10 minutes) between sunset times in the north and south, and so we do not anticipate substantial differences between these locations.

Number	**Reviewer comment**	**Author response**
		At the Autumn transition, clocks go back 1 hour, resulting in an hour more sleep - In the morning, civil twilight sunrise times change from approximately 7-7:30am to 6-6:30am across Great Britain. There is an hour more sleep and mornings are lighter by an hour before the morning peak travel time, therefore compounding the sleep and light hypotheses and resulting in the most appropriate conditions for a reduction in casualties. In terms of regional impacts, the most eastern locations experience civil twilight sunrise approximately 23 minutes before the most western locations, and so eastern locations are expected to report greater reductions in casualties. Furthermore, the most southern locations experience sunrise approximately 21 minutes before the most northern locations and so southern locations are expected to report greater reductions in casualties. - Civil daylight occurs throughout the inter-peak period (10-4pm), and so there are no anticipated effects from the light hypothesis. However, the sleep hypothesis could apply, as there is an extra hour of sleep gained throughout the day, potentially leading to a reduction in casualties. - In the evenings, civil sunset times change from approximately 6-6:30pm to 5-5:30pm across Great Britain. There is an hour more sleep throughout the day but evenings are darker by an hour during the

		PM peak of traffic. In this situation, there is a potential conflict between the sleep and light hypotheses. In terms of regional effects, sunset in the most eastern locations occurs approximately 24 minutes before sunset in the most western locations, and so we can expect more casualties in the east. Sunset in the most northern locations occurs approximately 13 minutes before sunset in the most southern locations. Though it is difficult to ascertain whether it is feasible to expect regional differences, it could be plausible to anticipate more casualties in the north. We have updated the discussion in Section 4 to explicitly state the above reasoning. Unfortunately, it is not possible to disentangle the effects of the two hypotheses when they are in conflict in our analysis, and so we have mentioned this as a limitation in Section 4.4.
9	You test the light hypothesis by looking for changes in road casualties in the PM peak and night time periods. However, I'm not sure we would expect any effect in the night time period as this period is in darkness both before and after the clock change - there is little to no difference in ambient light before and after.	The PM peak period is defined from 16:00 – 18:59 and the night period is defined from 19:00 – 23:59. It is true that in the Autumn transition, the sunset time change occurs within the PM peak period, with sunset changing from approximately 6-6.30pm to 5-5.30pm across Great Britain. However, in the Spring transition, the sunset time changes from approximately 7-7.30pm to 8-8.30pm across Great Britain and this occurs within the night period. We have added this clarification in Section 4.

Number	Reviewer comment	Author response
10	To assess both light and sleep hypotheses it seems to me that you would need to compare changes in casualties in the relevant period (e.g. PM peak for light, AM peak for sleep) with changes in other periods, to account for any changes in casualty numbers due to other factors unrelated to these two hypotheses (for example, school vacation periods could consistently fall on one side of	If we had designed the analysis to investigate changes in accident event *severity* at the transitions i.e. casualties per daily traffic volumes, then it would be important to disentangle the effect of potential traffic volume changes at the transition, but this is not the case in our analysis. In our analysis, we have defined the outcome measure as the *absolute* number of total casualties/fatalities as we are interested in

	the clock change, potentially influencing casualty rates due to increases (or decreases) in road traffic. Equally, weather conditions could be another conceivable factor, with potentially worse weather conditions falling before the Spring clock change and after the Autumn clock change, which could offset changes in road casualties due to the light and sleep hypotheses).	absolute increases or reductions in total casualties/fatalities at the transition. As such, we anticipate that the average treatment effect would pick up changes that represent the compound effect of changes in absolute traffic volumes and changes in casualties due to the transition, and this is deemed acceptable. As we wish to ascertain whether there are changes in total casualties and fatalities at the transition, it does not matter if this is attributed to the fact that there is a discontinuity in traffic volumes at the cutoff as a result of daylight savings transition effects. We highlight that while it is acceptable for our model to pick up traffic volume changes due to transition effects, we must still condition out exogeneous changes in traffic volumes which cannot be attributed to the transition, and this is the reason why we include seasonal covariates for year, day of week, and time period in the models. In terms of school holidays, unfortunately we were not able to obtain a data set that identifies every school holiday in each local council over the 14 years worth of data used in our models. We acknowledge that this could potentially influence the results, and we have included this as a limitation of our analysis in Section 4.4. In terms of weather, again, data at a time period level for local area zone since 2005 was not available, and we have acknowledged this as a potential limitation. However, we would like to point out that the bandwidths are quite narrow around the transition (please refer to response to comment 1, the mean bandwidth is 4.3 days either side of the transition), and the narrow windows would minimise the degree of systematic impacts from school holidays and weather effects.. We have included discussion of the above points in Section 4.4.
	Minor comments:	-
11	Check spelling of “casual” throughout the paper	We thank the reviewer for picking this up and we have now amended the typos on pages 3 and 12.
12	Line 24-26, page 4 - there are a few recent studies that would support a causal link between changes in light conditions and road collisions, e.g.	We thank the reviewer for the references, and have added these to the introduction in Section 1.

	Uttley, J., & Fotios, S. (2017). The effect of ambient light condition on road traffic collisions involving pedestrians on pedestrian crossings. Accident Analysis & Prevention, 108, 189-200. Fotios, S., Robbins, C. J., & Uttley, J. (2021). A comparison of approaches for investigating the impact of ambient light on road traffic collisions. Lighting Research & Technology, 53(3), 249- 261. Raynham, P., Unwin, J., Khazova, M., & Tolia, S. (2020). The role of lighting in road traffic collisions. Lighting Research & Technology, 52(4), 485-494.	
13	"Collisions" or "Crashes" is preferred to "Accidents", e.g. see Davis, R. M., & Pless, B. (2001). BMJ bans "accidents": Accidents are not unpredictable.	We thank the reviewer for letting us know about this. We have now removed all generic uses of the term "accident" and instead refer directly to "casualties", except in cases where we refer to the data set. The UK Department for Transport uses the term "accidents" in their road traffic data set, so we have decided to retain the term when referring to the data set for accuracy and clarity in Section 2.2. We have included a definition to clarify the terminology, i.e. "Casualties are defined as personal injuries of any severity as a result of an accident event. As specified in [21], a single accident event can be associated with more than one casualty".

Number	**Reviewer comment**	**Author response**
14	Can the authors clarify what is meant by 'local bandwidth', in the context of the regression discontinuity method, particularly the "optimised local bandwidth" as used in line 47, page 4? In particular, why different bandwidths are used for different models? This could raise questions about consistency and 'cherry picking' of results to fit with requirements.	As mentioned in the previous response to comment 1, the concept of selecting the optimal bandwidth based on minimising the mean squared error (MSE) of the average treatment effect is a well-established concept in the RDD literature (Cattanaeo and Vazquez-Bare, 2016; Calanico et al., 2015; Imbens and Kalyanaraman, 2012). This selection procedure is data-driven and objective, and selects the shortest (i.e. local) bandwidth in the vicinity of the cutoff subject to the minimisation of the MSE, thus ensuring that the key assumption of random treatment is upheld, which in turn enables a casual interpretation of

			the model results. The optimal bandwidth selection process is considered to be objective (i.e. no issues with cherry-picking) and less prone to bias when compared to nominating an arbitrary bandwidth as was common in the earliest implementations of RDD (Cattanaeo and Vazquez-Bare, 2016). For the benefit of a general readership who may not be familiar with RDD methods, we have included the aforementioned explanation in Section 2.2.
	15	Line 48, page 4 - Can you say more about what was missing from these 0.02% of observations?	These records that were removed were those that did not have spatial information (i.e. latitudes or longitude values) or timestamp records. Without this information, we are not able to allocate the time period or spatial band information required for the models, and so these records had to be removed. Clarification of this has been added in Section 2.1.
	16	Can you explain why Bank Holiday records were removed?	These records were removed as Bank holidays could potentially represent abnormal out-of-season traffic levels which could confound the baseline time trends before and after the DST transitions. Clarification of this has been added in Section 2.1.
	17	Can the code and data for the analysis be made openly accessible?	The data is open-source data from the Department for Transport, and we have provided the link to the data set in the "Data availability statement" at the end of the paper. We have now also provided the R code for the models.
Reviewer 3 -		***Prof. JA Hanley, McGill University***	
	1	The normal error model has advantages over Poisson etc, in that the variances will be more accurately measured (and not model-based) But you do have small counts with the time sliced so finely. any way to have a falsification test? maybe slide time by 2 weeks and pretend DST kicked in earlier later	We thank the reviewer for the suggestion of performing placebo testing at a cutoff date other than the daylight savings transition (DST) dates. Rather than using an arbitrarily defined alternative cutoff date, we follow the advice of Imbens and Lemieux (2008) in their seminal paper on RDD modelling. As per Imbens and Lemieux (2008), we partition the data to obtain two smaller datasets before and after the DST transition dates. We then nominate a placebo cutoff date in each dataset, which is equivalent to the mean value of the running variable in each dataset. Therefore, we undertake two placebo tests for every original model: 1. we perform a RDD with a cutoff before DST, and 2. we perform a RDD with a cutoff after DST. The majority of the significant models have passed the additional placebo

		specification tests, however, 4 have failed and have therefore been removed from the final set of significant models reported. The overall findings of the paper remain largely unaltered; the DST transitions have a minimal impact on road casualties. The placebo testing has been detailed in the paper in Section 2.3.

Number	Reviewer comment	Author response
2	show middle of the day results separately, and show all data graphically -- raw as well as fitted models. and by period of day...	All time of day results, including the middle of the day results (termed the "inter-peak" period, coded number 3) have been provided in Tables 3 and 4. We have now included an additional note with the two tables to make clear which coded numbers refer to which time of day period. In terms of a graphical representation of all significant models, we have included a full set of these with the reviewer response. The plots include the raw observations ("observations"), the fitted values ("fitted_values"), and the time trends excluding the fixed effects for seasonal variation ("pre_dst_trend" and "post_dst_trend"). We have generated two plots for each scenario: the first shows all data points including the extent of raw observations, and the second is zoomed in to highlight the time trends. We are of the opinion that the plots themselves do not provide additional valuable information, as the main output of the RDD modelling is the magnitude and significance of the average treatment effect, which has been provided in Tables 2, 3, and 4. However, if the reviewer is still of the opinion that the plots should be included, we are happy to provide these in an Appendix to the main paper.
3	readers need to understand why in Spring, raw counts are higher post than pre, but model fit suggests otherwise.. seeing the data is more likely to help convince readers that what you did/found is reasonable.	In RDD modelling, the average treatment effect quantifies the impact of an intervention in the immediate vicinity of the cutoff. By definition, the window around the cutoff must be sufficiently short so that the key assumption of random treatment is upheld. As mentioned in our response to Reviewer 2's comment 1, the initial raw data counts are for a wider 3 week period either side of the transition dates so that we have enough data points to calculate the local optimal bandwidth. Local optimised

		bandwidths are calculated for each model, and after calculation, the maximum window either side of the transition dates is 7.8 days, the minimum is 1.8 days, and the mean is 4.4 days. This changes the before/after casualty counts in each model, and is the reason why we observe reductions or no significant effects at the Spring transition in the RDD models, compared to what the original 3 week time periods suggested. In Tables 2, 3, and 4, we have now included the casualty numbers before and after the cutoff for each model to make this clearer, and we have added an accompanying explanation in Sections 2.1 and 3.1.
4	Disc. of USA results (where FARS system is more complete) -- and any other countries also, would help.	In Section 1, we refer to a study by Smith (2016) who reports an increase in road casualties at the DST transitions in the USA. Two other studies in the USA by Coate and Markowitz (2004) and Crawley (2012) report reductions in casualties. Furthermore, in Germany, Lindenberger et al. (2019) report no significant impacts in their analysis of road casualties, and Robb and Barnes (2018) report increases in road casualties in New Zealand at DST transitions.

Number	Reviewer comment	Author response
5	Giving more of the coefficients of the fitted models would also help us judge if fits seem sensible and accord with what would be expected.	The outcome of interest in a regression discontinuity design (RDD) model is the magnitude and significance of the average treatment effect, which we have reported in Tables 2, 3, and 4. The other terms in the model are exogeneous conditioning terms for seasonal variation, and do not provide valuable information on what happens at the daylight savings transition cutoff point. Moreover, since the basis of RDD is to analyse local behaviour very close to the threshold, it is very difficult to define an expectation for the magnitude and sign of the coefficients associated with the conditioning terms in the model. The intercept, the localised before and after time trends, and the covariates that represent year and the day of the week could equally be positive or negative and of varying magnitude in each spatio-temporal sample. As a result, it is not conventional to report on the additional covariate

		coefficients in RDD literature. However, we have extracted all coefficient values for all significant models and have summarised these in the tables attached. If the reviewer is of the view that these values should be included in the paper, we are happy to provide these in an Appendix (though please bear in mind that the accompanying interpretation of these can only be very limited).
6	any place for a finely matched analysis (e.g. Mantel-Haenszel, with difference in differences)?	A matched analysis would be possible if there were local zones in Great Britain that did not undergo the daylight savings transition. However, all temporal and spatial samples in our analysis underwent the transition, and are unambiguously considered to be "treated" after the transition date, and so matched analyses are not possible in this case.
7	What happens if do this for July or September or May? (gives some sense of variations under the null)	We believe that this point has been covered by undertaking two additional placebo tests for each original model before and after the DST dates as per the recommendations in Imbens and Lemieux (2008) (refer to response to comment 1). The placebo tests indicate that the majority of models with significant average treatment effects are correctly specified, however, 4 models did not pass the tests and have therefore been removed from Tables 3 and 4. Further detail on the placebo tests have been provided in Section 2.3.
8	Are you over-selling 'causal'. What makes yours more causal than others? It is still regression, and using regression models to turn back the clock ('setting' things this week to what they 'would have been') is easier said that done.	We agree with the reviewer that the terminology "casual" is not ideal, however, this is the terminology adopted in the statistics literature so we retain this in the paper. In the statistics literature, the regression discontinuity design (RDD) framework is defined as a statistical technique that quantifies the causal impact of a given intervention on a given outcome measure. It is deemed that causal inferences are able to be made using this framework due to the stringent assumptions that must be met when undertaking the analysis, which we have detailed in Sections 2.2 and 2.3. We have directed readers who may not be familiar with the method and terminology to seminal papers on RDD including Imbens and Lemieux (2008) and Lee and Lemieux (2010) in Section 2.2.

	The split b/w light and sleep is a plus.	We thank the reviewer for this comment.

VERSION 2 – REVIEW

REVIEWER	Hanley, JA McGill University, Department of Epidemiology and Biostatistics
REVIEW RETURNED	04-Jan-2022

GENERAL COMMENTS	Dear Editots and Authors Last time I suggested 2 improvements. 1. Explain to readers,and especially skeptics, how your modelling 'reverses' the trend seen in the raw data inTable 1. 2 Present plots of the raw data, with the fitted models overlaid on them. This is a matter of transparency, something that is vital in a charged topic such as this. Imagine talking to a journalist who has to explain all the models to the public, and to politicians. They could not possibly follow all the technical details, and so they will want other props and intuition. These pleas of mine seem to have gone unheeded. A general to the journal Editor. What's with this obsession of journals to take up half of the title with a description of the methods, rather than the question being addressed. When Nature published the paper by Watson and Crick, did it use a title such as "An X-ray-diffraction study of the structure of DNA" ?

VERSION 2 – AUTHOR RESPONSE

Reviewer 3 - Prof. JA Hanley, McGill University

Last time I suggested 2 improvements.

Explain to readers,and especially skeptics, how your modelling 'reverses' the trend seen in the raw data inTable 1.

- We understand the concerns of the reviewer, and so we have now clarified this further by adding the following text in Section 2.1: "... it should be noted that this study considers the impact on casualties in the immediate vicinity of the transition dates, and so the 3 week windows will shrink considerably after calculation of the optimal bandwidth around the transition dates for each model. Therefore, the general trend for the aggregate 3 week windows showing more casualties after the transitions may not be applicable at shorter bandwidths."

Present plots of the raw data, with the fitted models overlaid on them. This is a matter of transparency, something that is vital in a charged topic such as this.

Imagine talking to a journalist who has to explain all the models to the public, and to politicians. They could not possibly follow all the technical details, and so they will want other props and intuition.

These pleas of mine seem to have gone unheeded.

- We do appreciate that including the raw plots would be helpful for non-technical audiences, and so we have now included the plots as supplementary material, and have included a brief description of the plots in Section 3 as follows: "For further information, we have additionally included plots for every significant model including the original observations and fitted values as a Supplementary file. We have generated two plots for each scenario: the first shows all data points including the extent of raw observations, and the second is zoomed in to highlight the time trends".

A general to the journal Editor. What's with this obsession of journals to take up half of the title with a description of the methods, rather than the question being addressed.

When Nature published the paper by Watson and Crick, did it use a title such as "An X-ray-diffraction study of the structure of DNA" ?

- As the editor suggested, we have now changed the title from "A causal regression discontinuity design analysis of road traffic casualties in Great Britain at daylight savings time transitions", to "Road traffic casualties in Great Britain at daylight savings time transitions: a causal regression discontinuity design analysis".