Dear editorial board members and six reviewers,

Thank you very much for the opportunity to resubmit our manuscript, now titled, “High Profile European Football Matches Are Linked to Traffic Accidents in East Asia (BMJ-2020-054280.R1),” for further consideration for publication at the British Medical Journal. At the outset, we would like to thank each of you for taking the time to review this paper and for providing such constructive and helpful feedback. We very much appreciate all of the thoughts, comments, and questions you provided. Below is an executive summary of the major changes:

We have collected more data to replicate our findings from the Singapore data. Specifically, we were able to obtain daily traffic accident records from Taiwan between 2013 and 2018 (N = 1,814,320 accidents). As a result we also coded for more game days in those years. We are happy to report that all of our findings replicated. The Singapore and Taiwan data complement each other, as the former contain fine-grained data on driver and accident characteristics whereas the latter do not but contain many more accidents across both rural and urban areas. These new findings also show that our results are replicable beyond Singapore and taxi drivers.

We have collected survey data from taxi drivers (N = 100) and non-taxi drivers (N = 100) in Singapore to address two key issues – a) whether Singapore taxi drivers stay up late to watch European football games and b) whether Singapore taxi drivers’ football viewing habits are significantly different than Singaporeans who are not taxi drivers. We found that more than 1 out of 3 drivers indicated that they did stay up late to watch games and this pattern is similar among non-drivers. These new data also showed that people are most interested in watching football matches involving high market-value teams, which is an important assumption of our measure.

We conducted additional supplementary analyses to include games that involved at least one top 10 teams in any given league (all leagues have 20 teams). This is due to the fact that games involving two bottom-ranked teams might not be popular in Asia. We are happy to report that our findings replicated in both datasets.

We have significantly restructured our paper by separating the methods/analytical plan and results sections. We now include a detailed methods section to walk readers through our materials, coding, as well as analytical plan.

Please find below our point-by-point response to each of the comments made by the editorial board and the six reviewers. In our responses, we summarize our solutions to the issues raised and provide page numbers to specifically indicate where each issue was resolved in the manuscript. Adding these new data and details necessitates a longer paper, but we are more than happy to trim at a later stage.

Again, we really appreciate your great suggestions and guidance!

Yours Sincerely,
The authors

Editorial Board

* We are interested in the research question, which is the sort of slightly unusual angle that makes it a good fit for our Christmas issue.

We are glad to hear that this is a good fit for the Christmas issue!

* Papers published in the Christmas issue of The BMJ have to meet our methodological and reporting standards, however, even if they address an unusual or quirky research question. In its present form, this paper falls short. We'd like to give you the opportunity to address these shortcomings, in view of the fact that we have quite a bit of time before the Christmas issue. We can't make guarantees of publication, however. A revised paper will still need to be scrutinised by our statistician and it is possible he or we will not feel satisfied with the revised paper.

We appreciate the thorough attention that the review team is giving to our work, as well as the opportunity to revise our paper.

* Many editors commented that the study is poorly reported. Methods are interspersed in the results and introduction sections, and there is no discrete methods section in the paper.

We apologize for combining our methods and results sections. It was an attempt to make the paper more concise. In this revision, we have separated our methods and result sections. In the methods section, we use a step-wise approach to walk readers through our materials and methods:

First, we discuss how the traffic accident data from Taiwan and Singapore were obtained and their respective characteristics.
Second, we discuss clearly the coding of the football data. We provide reasons why combined team salary cap of a game is used as a proxy for the game's popularity as well as other details about our coding.
Third, we discuss, in details, the three sets of analyses we conducted. Specifically, the first analysis examines whether the average market value (i.e., popularity of games) on day k has a positive effect on traffic accident on day k. This is the broadest level of analysis that documents the basis of our proposed effect. The second analysis examines day-time vs. night-time accidents to rule out other accounts of our findings beyond sleep deprivation. The third analysis uses time-series analyses to rule out the possibility that average market value and number of traffic accidents were related because of an underlying temporal trend (e.g. both factors increasing linearly over time) or autocorrelated residuals.
Fourth, we discuss the behavioral survey study we conducted with Singaporean taxi drivers and non-drivers. In this description, we are specific in terms of the methods of data collection and questions asked in the survey.

Finally, we wish to note that all syntax to reproduce our analyses are provided in the OSF link we provided in the text (p. 2). We are happy to provide additional details and/or clarifications should the statistical advisor has any questions related to our analyses/syntax. We hope this structure is now much easier to navigate.

* Team market value seems like a very crude proxy for "popular games". A game between a top team and the team in last place in the league may be more popular than one between two middle of the table teams, who may accrue a higher market value (take Wolves vs Arsenal for example). We can't see Asian fans staying awake to watch Liverpool vs Norwich City... In fact, only a handful of matches are really high stakes and we wonder why you did not focus on those: derbys, matches that decide national championships, Champions League or Europa League (it is not Europa Champion League!) finals...

Thank you for this comment. We took this comment very seriously and addressed it with two approaches.

First, we wanted to get a better sense of what constitutes “popular games” for Singaporean taxi drivers in our sample. To do so, we surveyed 100 taxi drivers in Singapore while they were waiting for customers at taxi stands. We compensated all participants with SGD 5 (~GBP 2.8) for a couple minutes of their time. To avoid selection biases to the best of our ability, we recruited all participants at both day and night time.

In the survey, we asked them how likely (1 = very unlikely to 7 = very likely) they would watch a football game between: 1) a top vs. a bottom team (M = 1.85, SD = 1.23), 2) a bottom team vs. a similarly ranked bottom team (M = 1.67, SD = 1.09), 3) a top vs. a top team (M = 2.46, SD = 1.84), 4) their favourite team vs. a bottom team (M = 3.08, SD = 2.25), 5) their favourite team vs. a top team (M = 3.56, SD = 2.48), and 6) their favourite team vs. any team (M = 3.31, SD = 2.34). As you mentioned, we assume that an Asian taxi driver’s favourite team would very likely to be a top team. Unsurprisingly, viewership is highest between one’s favourite team vs. another top team, and this is also higher than a match between just any two top teams ($t [99] = 5.94$, $p < .001$). Moreover, participants indicated that they were more likely to watch a game played between their favourite team vs. a bottom team than a game played between two top teams ($t [99] = 3.11$, $p = .002$).

Armed with this insight, we decided to conduct supplementary analyses. In these supplementary analyses, we removed games that were played between two bottom teams of any leagues. We defined bottom teams as teams that were ranked between #11-20 (because all five leagues have 20 teams each) in all of the Big-Five Leagues at the end of the particular season. We updated the list of top-10 team annually because some bottom teams are relegated to a lower division. In other words, the games analyzed in the supplementary analyses were played by at least one top-10 team in any given Big-Five League. For your information, top-10 teams in the English Premier League consistently include Manchester United, Manchester City, Chelsea, Arsenal, Liverpool, Tottenham Hotspur, Liverpool, Everton, etc., all prototypical strong teams in the English Premier League. With this new coding, we are glad to report that all of our analyses replicated, with very similar effect sizes. These analyses can be found on pp. 30-31 in the SI.

We did not code for games from only derbies and matches that decide national championships for a few reasons. First, while most derbies are popular, other games could be equally if not more popular (e.g., a top team vs. another top team but not considered as derby). For example, the London Derby between Arsenal and Chelsea is no doubt popular, but it is hard to argue that it is more popular than a game between Manchester United and Arsenal (which is not a derby). We think that dropping these popular but non-derby games

would reduce the generalizability of our results. Second, some derbies are not publicly or formally recognized in Asia. For example, the West London Derby between Chelsea vs. Fulham is not often known as a derby for Asian fans. More generally, we suggest that derbies are often more popular among locals (because of the history of the rivalry) than international viewers (who likely only care about current top teams). Third, it is difficult to determine which games decide the national championships because the Big-Five Leagues all use a scoring system that accounts for every game in the season. Unlike American Football, for example, there is not a single game (i.e., Super Bowl) that determines the champion of a league. End-of-season games (games in May) might be more popular, but we account for this by the month-of-the-year control variable.

Finally, we wish to clarify that we did include the Champion League and Europa League games from the Round of 16 games till the finals. We clarified this on p. 6.

* Some aspects of the paper seems quite speculative. For example, you do not provide evidence that taxi drivers in Singapore are committed watchers of European football. We would want to see viewership numbers, to see if more watched games through the middle of the night were associated with more accidents later that day (presumably because people are driving around tired).

As mentioned in our previous response, we conducted a survey study with 100 random taxi drivers in Singapore. In addition to the results reported above, we also asked them “how many nights have you stayed up late to watch a European football game in the past month?” (0 = zero to 4 = four or more nights). We conducted this survey in mid-February, immediately after we received this revision request. This is important because mid-February is approximately the mid-season point for all of the Big-Five Leagues, as all of them started in August 2019 and will end in mid- or late-May 2020. If we had conducted the survey towards the end of the season, in May, self-reported viewership might have been inflated because this is the most popular time of the season.

We found that 37.4% of all drivers (1 did not respond) indicated that they had stayed up at least one night in the past month to watch European football games (1 night = 11.1%, 2 nights = 6.1%, 3 nights = 6.1%, 4 nights or more = 14.1%). All in all, more than 1 in 3 taxi drivers are regular late-night/early-morning football viewers, suggesting that this sample is relatively committed watchers of European football. Moreover, there was no significant difference between the average number of nights that taxi drivers ($M = .98$, $SE = .15$) and the general public ($N = 100$; $M = .70$, $SE = .12$, $p = .144$) stayed up late to watch games. The lack of difference was even more apparent when gender is controlled for ($p = .991$).

We attempted to obtain viewership data for each game in our data set in Singapore, but this is simply not possible due to a lack of publicly available television data, because all European football games are aired based on a paid subscription basis in Singapore (by private companies). Even if such data are available, we suggest that it would not be accurate because many people use illegal streaming services to view games. This is the reason why we used market value as a proxy for games’ popularity. We hope our survey with taxi drivers can assuage your concerns on this issue.

The modeling is difficult to judge given the absence of a statistical methods section, although some things are described in the results section. One editor commented that "looking at the graphs doesn't give me much confidence in the linearity. I would want more information on things like severity of accidents."

We again apologize for the lack of clarity about our methods and results section. As mentioned earlier, we now use a four-step approach to discuss our methods and analytical plan. We also separate our methods and results section to increase clarity. We also wish to note that all syntax to reproduce our analyses are provided via the OSF link listed in the text (p. 2). We are happy to provide additional details and/or clarifications should the statistical advisor has any questions related to our analyses/syntax.

Unfortunately, severity of accidents is not available in our Singapore data. However, we did obtain a second data set in Taiwan to replicate our findings more generally. We searched extensively online from all governments within the same time zone as Singapore. Many of the South East Asian countries, however, do not keep good public records (or any public records at all). Hong Kong does have publicly available traffic accidents, but only at the monthly level. At the end, we were able to secure a data set from the Taiwan government that contains all daily traffic accident across Taiwan from 2013-2018 (2019 numbers are not yet available). We conducted the same analyses and are happy to report that all results replicated in this new Taiwan data.

We now include both datasets in the paper, and they each have strengths and weakness.

Strengths of the Singapore data:
Contain rich driver demographics (e.g., age, driving experience, gender, etc.) that we can control for.
Contain fine-grained weather data, because the weather data were recorded specifically at the exact time and location (street-level) of the accident.
Weaknesses of the Singapore data:
Contain data from 2012-2014 (three years; N = 41,538 accidents).
Contain data only among taxi drivers, although we did conduct a survey study and found that a good number of taxi drivers watch late-night European games regularly and that their football viewing habits are not that different compared to non-drivers.

Strengths of the Taiwan data:
Contain data from 2013-2018 (6 years and more recent; N = 1,814,320 accidents).
Contain data from among all drivers in Taiwan, in both rural and urban areas.
Weaknesses of the Taiwan data:
Does not contain driver demographics, because this is a government data set.
Does not contain fine-grained weather data. We attempted to code for weather data by using the Taiwan Central Weather Bureau’s records. Unfortunately, we were unable to code for hour-level and street-level weather data. The Taiwan Central Weather Bureau only has data on whether it rained on a particular day in a particular city. Using this weather record will be highly inaccurate because accidents happened at different hours of the day and different locations within a city.

As you can see, although neither dataset is perfect, they do complement each other very well. We are glad to report that our findings are supported in both datasets, suggesting that our findings are replicable beyond Singapore and possibly to other cities within the GMT + 8 time zone. Nevertheless, we did include a sentence in the discussion section to discuss the limitation of being unable to discern the total human causalities as a result of our findings (p. 13).

Finally, we now control for the quadratic function in both the Singapore and Taiwan data. We are happy to report that the linear function remains statistically significant in both data sets, while results on the quadratic function are inconsistent. These statistics can be found in the SI (pp. 29-30).

* Our statistical consultant commented that he can't follow the methods, or what the control days are that were used for comparison.

We hope that our new methods section (which includes an extensive "analytic plan" section) is easier to follow. We have also annotated our online code in order to make our analyses more accessible.

There are no purely "experimental days" or "control days" in our analysis. Rather, we use a continuous proxy for the popularity of football matches (average market value) to compare traffic accidents on days with relatively more popular football matches to accidents on days with relatively less popular football matches. Our new behavioral survey and supplemental analyses suggest that this market value proxy is a valid and reliable indicator of football match popularity.

Hopefully our revised text is easier to follow.

* Is the information about cars and drivers specific to the taxi cars?

That is correct. The color of the cars and demographics of the drivers are specific to the cars and drivers when the accidents occurred.

* You have not considered non-linear relationships between market value of the football match and the number of accidents.

As mentioned above, we now control for the quadratic function in both the Singapore and Taiwan data. We are happy to report that the linear function remains statistically significant in both data sets. Details can be found in the supplementary information on pp. 29-30. The quadratic effect is positive and statistically significant in Taiwan but negative and marginally significant in Singapore, suggesting that non-linear association is inconsistent with our overall data. We therefore avoid any strong claims about non-linear associations between high profile football matches and traffic accidents, and our overall data suggest that a linear relationship (as we have consistently demonstrated) is much more likely.

* How did you account for multiple football matches screened live on the same day?

On most game days there were multiple football matches. We took the average of all games' combined salary cap between the two games to form our independent variable. In the supplemental "top 10" analyses, we excluded games that did not contain one of the leagues' top 10 teams and averaged across games that contained at least one top 10 team.

* We found it unusual that there is exactly 1.00 extra accident for each increase in the mean market value of 170 million. On inspection the beta coefficient is 0.0002, which corresponds to a ratio of counts of 1 between 2 days where the match value differs by 1-unit. Hence, the variable does not seem important, but this is probably because a 1-unit increase is tiny compared to the millions of units increase. The value may be 1.005555, which when multiplied by millions has an impact. You need to clarify what's going on here.

The exact 1.00 value is an artifact of incidence rate format and the scientific norm to report two significance figures. The incidence rate corresponds to the percent change in the measured outcome (number of accidents) based on a 1-unit change in the predictor. An incidence rate of 1.00 would mean that the outcome is not changing at all based on the predictor. Our incidence rates (1.00015 in Taiwan and 1.00021 in Singapore) are quite low because the predictor (market value in millions) has a large range. Average market values are in the hundreds of millions, so a single million-dollar increase has a relatively low effect on traffic accident rate. However, as you note, when multiplied across hundreds of millions of dollars in average market value and millions of drivers, this still translates into many traffic accidents. So the apparently low incidence rate actually connotes a potentially dramatic impact.

We have tried to clarify the meaning of our terms in the revision. For example, we have expanded the incidence rate format in Table 1 to include more decimal points, so that all the incidence rates do not appear as "1.00." We have also explained in our table caption and the text that our incidence rate values translate to an additional 1 accident for every €8.99 million euros in the Taiwan dataset, and 1 accident for every €134.74-145.68 million euros in the Singapore dataset (depending on whether the model contains demographic information or not). These values are different because the datasets differ so much in size and predicted changes in Poisson models are sensitive to base rates (a weaker predictor can cause a 1-unit increase more easily if the outcome variable has a high base-rate). Finally, we have revised on impact analysis to incorporate our Taiwanese data, and make our findings as concrete as possible.

* It's not clear if there are any missing data and how this was handled

We did not analyze days where no football matches were played (over the mid-summer). These data-points were not included in our regression models. We have clarified this in our text on p. 6. There were no other cases of missing data (p. 7). We have clarified this in the main text.

* There is very strong causal language in many places, eg "Watching football (soccer) games from distant time zones (e.g., at 3am local time) increases the prevalence of local auto accidents"

We apologize for overstating our conclusions. As with most non-experimental research, we cannot claim definitive causality and as such we have removed all causal languages (e.g., replace “increases” with “is positively associated”).

* There are some unhelpful statements such as: “This increased rate of traffic accident may translate to between 382.12 and 8,182.44 accidents” –the width is wide and we don’t know the actual time-course here. Is this per day, per year, per season?

We revise our impact analyses and no longer include a range. We only estimate the number of accidents and economic impact due to taxi drivers in the Singapore data set (which likely is a conservative estimate) and estimate the number of accidents and economic impact due to all drivers in Taiwan. Please see p. 13 as well as pp. 35-36 in the SI.

* You model team capital against accidents. Would it not be more appropriate to look at top teams only?

This is a great point. As we discussed above, we conducted exactly such supplementary analyses. In these supplementary analyses, we remove games that were played between two bottom teams of any leagues. We defined bottom teams as teams that were ranked between #11-20 (because all five leagues have 20 teams each) in all of the Big-Five Leagues. In other words, the games analyzed in the supplementary analyses were played by at least one top-10 team. For your information, these teams in the English Premier League include Manchester United, Manchester City, Chelsea, Arsenal, Liverpool, Tottenham Hotspur, Liverpool, Everton, etc., all prototypical strong teams in the Premier League. With this new coding, we are glad to report that all of our analyses replicated, with very similar effect sizes. These analyses can be found on pp. 30-31 in the SI.

* Our patient editor commented that the paper is missing a statement of patient and public involvement and dissemination plans.

We definitively appreciate the importance of patient involvement in research. However, given the unique nature of our work we do not believe a statement of patient and public involvement is applicable because in this research there were no 1) “patients,” 2) recruitment of human subjects, and 3) active interventions. That said, we are more than happy to write such a state should the editorial board deem necessary.

We would like to close by thanking the entire editorial board for the very constructive comments. Thank you for pushing us to collect more data to clarify our assumptions (i.e., the survey study) and replicating our findings (i.e., the Taiwan data). We believe our paper has improved significantly as a result of them, and we are eager to hear your feedback again. Thank you!

Reviewer 1

It should be noted that it is an interesting research and an original idea. The basic assumption in this research is that high-profile football matches are of common interest in the Singapore's population and that a relatively large number of taxi drivers do follow them. This issue was not addressed in the manuscript and, I guess, can be easily added (look for Expedia poll from May 2019 stating that 3 out of 4 Singapore football fans plan their holidays around sporting events, for example).

Thank you for your kind word about our paper. That travel agency statistic is very interesting. We decided to take a more direct route to confirming that taxi drivers (and the general public) in Singapore are football fans. We conducted a primary survey study with 100 random taxi drivers in Singapore while they were waiting for customers at taxi stands. We compensated all participants with SGD 5 (~GBP 2.8) for a couple minutes of their time. To avoid selection biases to the best of our ability, we recruited all participants at both day and night time.

We asked them “how many nights have you stayed up late to watch a European football game in the past month?” (0 = zero to 4 = four or more nights). We conducted this survey in mid-February, immediately after we received this revision request. This is important because mid-February is approximately the mid-season point for all of the Big-Five Leagues, as all of them started in August 2019 and will end in mid- or late-May 2020. If we had conducted the survey towards the end of the season, in May, self-reported viewership might have been inflated because this is the most popular time of the season. We found that 37.4% of all drivers (1 did not respond) indicated that they had stayed up at least one night in the past month to watch European football games (1 night = 11.1%, 2 nights = 6.1%, 3 nights = 6.1%, 4 nights or more = 14.1%). All in all, more than 1 in 3 taxi drivers are regular late-night/early-morning football viewers, suggesting that this sample is relatively committed watchers of European football. This information is included on p. 9.

The second item that, in my opinion should be better elaborated is the definition of a car incident that has an ample range and, if data exists (or a proxy like the cost of fixing the damage), the whole work can be upgraded and maybe more précised conclusions can be drawn. In case it is not possible to get those data items, it should be also noted.

We agree that that severity of accident is something that is important. Unfortunately we do not have such data from our data which we note as a limitation on p. 13. However, we did obtain another data set from Taiwan to replicate our findings. We are happy to report that all results replicate.

Thank you for your time in reviewing this paper.
Reviewer 2

As a lay reviewer, I find it impossible to comment on the statistical evidence as I do not understand the processes used, nor the technical results they produce. The article also includes a high level of jargon which I imagine is only comprehensible to statisticians.

We have tried to minimize our use of jargon, but we do need to be precise in our language regarding the technical aspects of our analyses. However, as the Editor’s letter mentioned,

the BMJ editorial team for our submission includes a statistics expert to make sure we are on the right track.

The article cannot be assessed as a carer or a patient as it has no direct effect on those with long term health conditions and does not relate to any treatments.

We agree that our paper does not really relate to any treatments or focus on long term individual health outcomes. Fortunately, the Christmas issue of BMJ has enough latitude that our topic is within the mission of that issue. In addition, we do believe that these findings have some policy implications for the strategic scheduling of popular European football games (see our responses later). If public health officials can be considered “carers,” then our paper would have an important message for this population.

It would appear that the scheduling of football matches has an effect on the incidence of accidents during and after high profile games and that this is in direct proportion to the profile of the game in question.

Yes, this is the primary finding.

All data used for analysis has been obtained from just one source – a large taxi operator. This company’s cars clearly make up a large proportion of cars out on the road at any given time.

The author then goes on to report that when a high profile matches are shown, the accident rate goes up noticeably. The assumption is that this rise is directly related to drivers being tired, driving whilst listening to the match, or keeping an eye on the score. This implies that the taxi drivers are not only victims of these irresponsible drivers, but also perpetrators.

If accidents logged by all drivers were analysed there may well be a different picture. It would seem rather restrictive to focus on this one company, resulting in an unbalanced view.

These are great points. We also wish to have daily traffic accident data from all vehicles in Singapore. Unfortunately, such data are not available and we had to contact the largest taxi company in Singapore to obtain their traffic accident records. This is a limitation, which is why we conduct another study in Taiwan that does not rely on a specialized population. The fact that our results replicate so consistently in Taiwan suggests that we are documenting a general phenomenon, rather than one that is exclusive to taxi drivers in Singapore.

We now include both data sets in the paper, and they each have strengths and weakness:

Strengths of the Singapore data:
Contain rich driver demographics (e.g., age, driving experience, gender, etc.) that we can control for.
Contain fine-grained weather data, because the weather data were recorded specifically at the exact time and location (street-level) of the accident.
Weaknesses of the Singapore data:
Contain data from 2012-2014 (three years; N = 41,538 accidents).

Contain data only among taxi drivers, although we did conduct a survey study and found that a good number of taxi drivers watch late-night European games regularly and that their football viewing habits are not that different compared to non-drivers.

Strengths of the Taiwan data:
Contain data from 2013-2018 (6 years and more recent; N = 1,814,320 accidents).
Contain data from among all drivers in Taiwan, in both rural and urban areas.
Weaknesses of the Taiwan data:
Does not contain driver demographics, because this is a government data set.
Does not contain fine-grained weather data. We attempted to code for weather data by using the Taiwan Central Weather Bureau's records. Unfortunately, we were unable to code for hour-level and street-level weather data. The Taiwan Central Weather Bureau only has data on whether it rained on a particular day in a particular city. Using this weather record will be highly inaccurate because accidents happened at different hours of the day and different locations within a city.

As you can see, although neither dataset is perfect, they do complement each other very well. We are glad to report that our findings are supported in both data sets, suggesting that our findings are replicable beyond Singapore and possibly to other cities within the GMT + 8 time zone.

Also, if taxis are busier due to matches, then they are transporting passengers. Many of these will be listening to the match, but cannot be culpable for any accidents – it could be assumed that they could have commentary on whilst travelling which is distracting the driver.

Our second set of analysis partially ruled out this explanation. The data we have do not indicate that the accidents were caused by drivers who were watching a football game while they were driving. This is supported by the non-significant night-time accident effect in the Singapore data and the significantly smaller effect size for night-time accident, relative to day-time accident, in the Taiwan data. Given the time of day of the games (always late-night/early-morning before sunrise) and the times of days of the accidents, it is more reasonable to assume that the drivers were not watching the games while the accidents occurred, but rather were more likely to get into accidents because they were sleep deprived from staying up late to watch the games.

In conclusion, as the taxi company are recording higher than average accidents during high profile matches, then training of drivers and banning the use of radios, internet devices, etc. whilst driving would seem the first step in reducing the carnage.

We agree that it is probably wise to ban the use of any video-based equipment while driving, at least as a general policy. We mention this point on p. 14 when we discuss policy implications.

A taxi driver will spend lengthy periods of time driving and therefore tiredness must creep in at some point during their shift. This will add to the statistics, but is not necessarily as a direct result of the football.

Yes, shift length probably does increase the probability of an accident, but that is separate from the effect that we investigate. Unfortunately this variable is not available to us, but we did control for numerous other variables related to the drivers' characteristics.

There is a suggestion that kick off times could be brought forward from 7pm to 6pm. London is an hour behind the majority of European cities. Therefore this would result in a 5pm kick off in the UK – the time when most people are traveling home from work and would therefore either need to take time out of work or miss the match. This would have financial implications for both industry and the television companies.

We agree and have removed this recommendation. More generally, we suggest that strategically popular games on Friday and Saturday nights (local European time) may be a better solution. This is because most individuals in Asia do not have to work on the next morning (Saturday and Sunday mornings). As such, football fans won't be forced to drive to work tired and less likely to be exposed to tired drivers (p. 14).

International Sport and domestic sport is taking place all around the world at different times of the day and night depending upon residence. Whilst the point is made that Asia is a highly populated region, shifting the times of any sport will have an effect somewhere around the world.

As mentioned, we no longer recommend switching the start time of games, but rather strategically scheduling popular games on different days.

Overall, this research appears to be largely irrelevant as the source of information is questionable, and the limited possibility of solving the problem, other than stopping live streaming in order to play in full at a reasonable hour, questions its validity.

We have supplemented our findings with a separate independent source of information. This shows that our effects are replicable and generalizable beyond Singapore, and is likely true for many other cities within the GMT+8 time zone. We agree that designing an intervention to address the spikes in accidents that we have uncovered is difficult, because scheduling games at different times or days of the week can have economic implications. However, we believe that documenting the effect is an important first step that is necessary before other people who specialize in policies, regulations, and sports can dig further into the issue and work toward a solution. Ignoring the effect because we alone cannot solve it seems to be a less desirable option that documenting it.

Thank you for your time in reviewing this paper. We hope our responses and new data can assuage your concerns.
Reviewer 3

This is a well written paper and is easy to read and understand. The results are easily understandable and not published before. The authors have made clear they are not suggesting a causal relationship.

Thank you for your kind words. We appreciate the time you took in reviewing this paper.

I have some comments:
The data is 5 years old and merits and explanation re why so delayed.

We agree that the data might be seen as dated, but the first author was not aware of the existence of this data set until mid-2019. Prior to submission to BMJ, we did contact the taxi company to solicit more recent data. Unfortunately, our request was declined.

In this revision, we searched extensively online from all governments within the same time zone as Singapore. Many of the South East Asian countries, however, do not keep good public records (or any public records at all). Hong Kong does have publicly available traffic accidents, but only at the monthly level. At the end, we were able to secure a data set from the Taiwan government that contains all daily traffic accident across Taiwan from 2013-2018. This is the most up-to-date data available (2019 numbers are not yet available). We conducted the same analyses and are happy to report that all results replicated in this new Taiwan data.

We now include both datasets in the paper, and they each have strengths and weakness.

Strengths of the Singapore data:
Contain rich driver demographics (e.g., age, driving experience, gender, etc.) that we can control for.
Contain fine-grained weather data, because the weather data were recorded specifically at the exact time and location (street-level) of the accident.
Weaknesses of the Singapore data:
Contain data from 2012-2014 (three years; N = 41,538 accidents).
Contain data only among taxi drivers, although we did conduct a survey study and found that a good number of taxi drivers watch late-night European games regularly and that their football viewing habits are not that different compared to non-drivers.

Strengths of the Taiwan data:
Contain data from 2013-2018 (6 years and more recent; N = 1,814,320 accidents).
Contain data from among all drivers in Taiwan, in both rural and urban areas.
Weaknesses of the Taiwan data:
Does not contain driver demographics, because this is a government data set.
Does not contain fine-grained weather data. We attempted to code for weather data by using the Taiwan Central Weather Bureau's records. Unfortunately, we were unable to code for hour-level and street-level weather data. The Taiwan Central Weather Bureau only has data on whether it rained on a particular day in a particular city. Using this weather record will be highly inaccurate because accidents happened at different hours of the day and different locations within a city.

As you can see, although neither dataset is perfect, they do complement each other very well. We are glad to report that our findings are supported in both datasets, suggesting that our findings are replicable beyond Singapore and taxi drivers.

The other 'variables' /associations could have played apart and were these careful statistically analysed? Age, gender, total duty hours.

Yes, drivers' demographics including age, gender, race, educational level, and driving experience were controlled for in the Singapore data set. In addition, we also controlled for weather data in the Singapore data. These can be found on p. 10. Unfortunately, the data set did not contain total duty hours.

Were the drivers asked if they were fans of football/sleep deprivation 9and its cause), excitement and if they watched the matches?

We did not have direct access to taxi drivers. In this revision, however, we conducted a survey study with 100 random taxi drivers in Singapore. We also asked them "how many nights have you stayed up late to watch a European football game in the past month?" (0 = zero to 4 = four or more nights). We conducted this survey in mid-February, immediately after we received this revision request. This is important because mid-February is approximately the mid-season point for all of the Big-Five Leagues, as all of them started in August 2019 and will end in mid- or late-May 2020. This is important because responses to this question will likely be inflated should we ask during the latter phases of the season. We found that 37.4% of all drivers (1 did not respond) indicated that they had stayed up at least one night in the past month to watch European football games (1 night = 11.1%, 2 nights = 6.1%, 3 nights = 6.1%, 4 nights or more = 14.1%). All in all, more than 1 in 3 taxi drivers are regular late-night/early-morning football viewers, resulting in sleep deprivation and suggesting that this sample is relatively committed watchers of European football.

Was there co-relation with previous accidents?

Although we have all drivers' characteristics related to each specific traffic accident, we do not know the identities of the drivers. In other words, there is no way to track whether the same drivers were in multiple accidents in a given year. On aggregate, however, the time series of traffic accidents has negative autoregressive processes, which suggests that traffic accidents on day t are followed by fewer traffic accidents on day t+1. This is not surprising, since a day with many traffic accidents will probably lead drivers to be more cautious the next day.

Any other cities looked at and if these can be reproduced?

This is a great question and as mentioned we are able to replicate our findings using data from Taiwan. We tried to search for other possible datasets, but as discussed, no other countries/cities within the GMT + 8 time zone keep daily traffic accident records.

There are many assumptions in the paper which need some explanation. It is possible that all these have been looked at already but not explicit in the submission.

We have re-organized our paper to make our assumptions clearer, and to show the reader how we supported these assumptions. This involved a lengthier method and analytic plan sections, and a behavioral study that explicitly confirmed two of our assumptions: (a) that Singapore taxi drivers are indeed football fans and (b) that market value is a good proxy for level of interest in football matches. Hopefully these added data and additional sections make our logic clearer.

Thank you for taking the time to review this paper and for pushing us to collect more data to replicate our model. We hope your find our revision responsive and convincing.

Reviewer 4

With great pleasure I read the paper entitled “Traffic Accidents as a Risk of Watching Football (Soccer) at Home: An Observational Study” by Yam et al. submitted for publication to the BMJ. This paper reports that watching high-profile football games in other time zones can be dangerous for roadside. The authors claim that this is especially problematic in Asia, because drivers lose sleep watching high profile games played in Europe which occur during local times in which they typically sleep, leading to a higher prevalence of traffic accidents. I really like the idea behind this study. However, I have several aspects that need further attention.

Thank you for your kind words and helpful review.

1.) Please give a more detailed overview which games you included. More football fans will watch games in the final phase of the tournaments. Also, tournaments will be more watched than national football league games. Furthermore, there are games in the final phase of a national football league that are more exciting and more widely watched (i.e. when a game decides who will be national champion). Finally, the best way to control for all these aspects is to have objective data on the number of TV-spectators of every game. Is it possible to get such data and control for that? It not, at least control to some degree on how “exciting” a game was.

We did not code for games from only matches that decide national championships or in the final phase of the season for a few reasons. It is difficult to determine which games decide the national championships because the Big-Five Leagues all use a scoring system that accounts for every game in the season. Unlike American Football, for example, there is not a single game (i.e., Super Bowl) that determines the champion of a league. End-of-season games (games in May) might be more popular, but we account for this by the month-of-the-year control variable.

That said, we do want to get a better sense of what constitutes “popular games” for taxi drivers in our sample. To do so, we surveyed 100 taxi drivers in Singapore while they were waiting for customers at taxi stands. We compensated all participants with SGD 5 (~GBP 2.8) for a couple minutes of their time. To avoid selection biases to the best of our ability, we recruited all participants at both day and night time.

In the survey, we asked them how likely (1 = very unlikely to 7 = very likely) they would watch a football game between: 1) a top vs. a bottom team ($M = 1.85$, $SD = 1.23$), 2) a bottom team vs. a similarly ranked bottom team ($M = 1.67$, $SD = 1.09$), 3) a top vs. a top team ($M = 2.46$, $SD = 1.84$), 4) their favourite team vs. a bottom team ($M = 3.08$, $SD = 2.25$), 5) their favourite team vs. a top team ($M = 3.56$, $SD = 2.48$), and 6) their favourite team vs. any team ($M = 3.31$, $SD = 2.34$). We assume that an Asian taxi driver’s favourite team would very likely to be a top team. Unsurprisingly, viewership is highest between one’s favourite team vs. another top team, and this is also higher than a match between just any two top teams ($t [99] = 5.94$, $p < .001$). Moreover, participants indicated that they were more likely to

watch a game played between their favourite team vs. a bottom team than a game played between two top teams (t [99] = 3.11, p = .002).

Armed with this insight, we decided to conduct supplementary analyses. In these supplementary analyses, we removed games that were played between two bottom teams of any leagues. We defined bottom teams as teams that were ranked between #11-20 (because all five leagues have 20 teams each) in all of the Big-Five Leagues at the end of the particular season. We updated the list of top-10 team annually because some bottom teams are relegated to a lower division. In other words, the games analyzed in the supplementary analyses were played by at least one top-10 team in any given Big-Five League. For your information, top-10 teams in the English Premier League consistently include Manchester United, Manchester City, Chelsea, Arsenal, Liverpool, Tottenham Hotspur, Liverpool, Everton, etc., all prototypical strong teams in the English Premier League. With this new coding, we are glad to report that all of our analyses replicated, with very similar effect sizes. These analyses can be found on pp. 30-31 in the SI.

We attempted to obtain viewership data for each game in our data set in Singapore, but this is simply not possible due to a lack of publicly available television data, because all European football games are aired based on a paid subscription basis in Singapore (by private companies). Even if such data are available, we suggest that it would not be accurate because many people use illegal streaming services to view games. This is the reason why we used market value as a proxy for games' popularity. We hope our survey with taxi drivers and the new data from Taiwan can assuage your concerns on this issue.

2.) Page 9, last line: 8,182.44: this number is based on the assumption that the general population is watching football as much as in taxi drivers. I would rather guess that taxi drivers watch football more often than the general population.

This is a great point. To address this issue, we surveyed 100 taxi drivers in Singapore while they were waiting for customers at taxi stands. We furthermore surveyed 100 random Singaporeans in two large malls as the comparison group. We compensated all participants with SGD 5 (~GBP 2.8) for a couple minutes of their time. To avoid selection biases to the best of our ability, we recruited all participants at both day and night time.

In the survey, we asked them "how many nights have you stayed up late to watch a European football game in the past month?" (0 = zero to 4 = four or more nights). We conducted this survey in mid-February, immediately after we received this revision request. This is important because mid-February is sort of the mid-season point for all of the Big-Five Leagues, as all of them started in August 2019 and will end in mid- or late-May 2020. This is important because responses to this question will likely be inflated should we ask during the latter phrases of the season. We found that 37.4% of all drivers (1 did not respond) indicated that they had stayed up at least one night in the past month to watch European football games (1 night = 11.1%, 2 nights = 6.1%, 3 nights = 6.1%, 4 nights or more = 14.1%). All in all, more than 1 in 3 taxi drivers are regular late-night/early-morning football viewers, suggesting that this sample is relatively committed watchers of European football.

Comparing between taxi drivers and general Singaporeans, there does seem to be a trend that the former stayed up more nights to watch football games than the latter (M = .98 nights,

SD = .15 vs. M = .70 nights, SD = 1.18, t = 1.47, p = .14). This association, however, is confounded by gender. In our taxi driver sample, 99% were male whereas in the Singaporean sample 50% were male. After controlling for gender, we found that taxi drivers and Singaporeans watched football games at about the same rate (t = -.01, p = .99). These results suggest that taxi drivers in Singapore are not necessarily more avid viewers of European football matches than male Singaporeans. More generally, this suggests that although our results were centered around taxi drivers, it is unlikely that this effect would be restricted to only taxi drivers. Indeed, even if it were only restricted to taxi drivers, people who drive cars which are not taxis still share the same roads as the taxi drivers, and are thus still at risk of being run into by taxi drivers.

We have also addressed this “general population” question by replicating all of our analyses using a large dataset of accidents in Taiwan. Since our results replicate in this general population dataset, it suggests that our findings generalize beyond taxi drivers.

3.) How can you exclude the possibility that taxi drivers are distracted by watching the game during driving? At least in my country (Europe) a lot of taxi drivers have TV screens, ipads, or can watch on their mobile phones. Of course, it is not allowed to watch during driving, but anyway, I think it is reasonable to assume that at least some of the taxi drivers are distracted by watching (or even listening) to matches while driving. I think this even more applies for all other drivers (not taxi drivers) as taxi drivers will likely be more adherent to the rule to not watch screens during driving. On the other hand, taxi drivers represent more than 5% of all vehicles during the night. Could you elaborate to this aspect?

Our second set of analysis partially ruled out this explanation. The data we have do not indicate that the accidents were caused by drivers who were watching a football game while they were driving. This is supported by the non-significant night-time accident effect in the Singapore data and the significantly smaller effect size for night-time accident, relative to day-time accident, in the Taiwan data. Given the time of day of the games (always late-night/early-morning before sunrise) and the times of days of the accidents, it is more reasonable to assume that the drivers were not watching the games while the accidents occurred, but rather were more likely to get into accidents because they were sleep deprived from staying up late to watch the games.

4.) Can you increase you data-pool? You now report on 13.000 taxi drivers of a single company. Do you have data from traffic accidents on a national level in Singapur? Or Uber? Or public transport drivers? Or increase the time period?

While we do not have additional access to other drivers in Singapore, we did obtain daily traffic accident records from the Taiwan Government as we mentioned above, which included a far larger pool of accidents (N = 1,814,320). Thank you for this comment.

5.) Usually, there are “hot-spots” in cities where more traffic accidents happen. So geographical differences exist with respect to where traffic accidents occur. Do you have data on that?

Unfortunately we do not have such fine-grained street-level data. Indeed, these data would compromise privacy.

6.) Why did you report your data in US Dollars? The clubs are in Europe, the original data of the market values are in Euros, and you submit this paper to an European journal. I see no value in reporting the values in US Dollars. One could make an argument for Pounds (as the journal is British), however, I see no reason to report US Dollars.

We agree and have converted all units into Euros.

7.) Finally, I don`t like the title, it does not read fluent. Please change.

The new title now reads “High Profile Football Matches in Europe Are Linked to Traffic Accidents in Asia.” We welcome your suggestions to further improve it.

Thank you very much for your constructive feedback. The paper has improved significant as a result of them!

Reviewer 5

Review of “Traffic Accidents as a risk of watching football at home: an observational study”

This study addresses an interesting hypothesis that watching European football games in Asian time zones increases traffic accidents in Singapore. While there is reason to justify this hypothesis, the study methods are insufficiently explained in the current draft and the discussion and conclusions are overstated given the limitations of the study design, which should be more thoroughly described.

As we note in our response to the Editor, we have restructured our paper to have separate methods and results sections. Also, in the process of making those revisions, we have added further detail to these sections. Moreover, we have added a more detailed discussion of the limitations of our analysis of the accidents in Singapore, plus added a supplemental study of accidents in Taiwan. That supplemental study has complementary strengths and weaknesses as the Singapore study. The Singapore and Taiwan data complement each other, as the former contain fine-grained data on driver and accident characteristics whereas the latter do not but contain many more accidents across both rural and urban areas. These new findings also show that our results are replicable beyond Singapore.

Throughout the paper the authors use causal language – “effect” of watching football, “increases” prevalence- all language should be associations rather than causal as this is not an experimental study. Overall this study presents a novel and interesting hypothesis, but would be improved from more critical discussion of the limitation of the data and avenues for future research rather than drawing conclusions about policy changes or determining that changing game times will improve the rate of traffic accidents (as this was not modeled and it is difficult to say whether that would make a difference).

You make a fair point that we did not conduct an experiment, and thus should avoid causal language. We have accordingly revised our language throughout the paper.

As we note in our response to your previous comment, we have also added a more detailed discussion of the limitations of our analysis of the accidents in Singapore, plus added a supplemental study of accidents in Taiwan.

As you recommend, rather than indicating that we have foolproof policy solutions for the problem that we have uncovered, we now discuss possible solutions as opportunities for future research.

Abstract:

The reviewer would not consider this to be a longitudinal study, as the study was not designed to collect multiple events from the same individual, but rather just included historical data over a few years.

Missing a word in the first sentence of the conclusions.

We no longer refer this as a longitudinal study and have corrected the sentence.

Introduction:
Is there evidence from TV networks about how many people are tuning in where? Or watching the matches lives vs recording them?

We attempted to obtain viewership data for each game in our data set in Singapore, but this is simply not possible due to a lack of publicly available television data (all European football games are aired based on a paid subscription basis in Singapore by private companies). Even if such data are available, we suggest that it would not be accurate because many people use illegal streaming services to view games. This is the reason why we used market value as a proxy for games' popularity. We hope our survey with taxi drivers and the new data from Taiwan can assuage your concerns on this issue. We describe these data thoroughly in our early responses to the editors.

Reason for choosing Singapore justified as affecting both those who stayed up late and those who did not (not sure why this is specific to Singapore)

We have removed this specific justification.

Information on the study population should be moved to methods section

We have done this.

Methods

There is no section labeled methods- add subheading. Also, there is insufficient information on modeling technique, processes for model selection and building, etc. Please add information on model selection.

We apologize for combining our methods and results sections. It was an attempt at brevity that in retrospect merely added confusion. In this revision, we have separated our methods

and result sections. In the methods section, we use a step-wise approach to walk readers through our materials and methods:

First, we discuss how the traffic accident data from Taiwan and Singapore were obtained and their respective characteristics.
Second, we discuss clearly the coding of the football data. We provide reasons why combined team salary cap of a game is used as a proxy for the game's popularity as well as other details about our coding.
Third, we discuss, in details, the three sets of analyses we conducted. Specifically, the first analysis examines whether the average market value (i.e., popularity of games) on day k has a positive effect on traffic accident on day k. This is the broadest level of analysis that documents the basis of our proposed effect. The second analysis examines day-time vs. night-time accidents to rule out other accounts of our findings beyond sleep deprivation. The third analysis uses time-series analyses to rule out the possibility that average market value and number of traffic accidents were related because of an underlying temporal trend (e.g. both factors increasing linearly over time) or autocorrelated residuals.
Fourth, we discuss the behavioral survey study we conducted with Singaporean taxi drivers and non-drivers. In this description, we are specific in terms of the methods of data collection and questions asked in the survey.

Finally, we wish to note that all syntax to reproduce our analyses are provided in the OSF link available in the text. We are happy to provide additional details and/or clarifications should the statistical advisor has any questions related to our analyses/syntax. We hope this structure is now much easier to navigate.

Do authors know how complete accident records are from the taxi company? Is there information on who is at fault from insurance claims, etc?

Our data contain all traffic accidents in that three-year period from the taxi company. We do have data on at-fault vs. not-at-fault accidents, and results were identical with this control. Because we don't have this data in the Taiwan data set, we do not include this control variable.

Results

Page 6: Generalizability refers normally to applying to broader populations. It is unclear how testing for interaction in this case increase generalizability.

We meant that the results are similarly strong across weekends and weekdays, and across rainy days and dry days, to give two examples. While "generalization" may be typically used in a medical setting to refer to populations, it is often frequently used to discuss the strength of results across different contexts. We originally used the word in this way—to show that our effects were similarly strong across different contexts. However, to avoid any confusion, we have removed the term when referring to our interaction tests.

Can authors stratify by key variables including weather and weekday/weekend, even though interaction was not significant?

This is an interesting idea, but it may subtract less than it adds. In cases with a non-significant interaction, it is considered bad practice to test for differences across groups because these differences may simply reflect sampling error than meaningful variation. For this reason, we have avoided stratified tests or simple slope probes. As an aside, our supplemental materials already contain 19 different modeling approaches which support the robustness of our findings (see SI), and we would prefer not to inundate readers with many more analyses, especially if they are not clearly helpful.

Discussion:

Important to call attention to the lack of individual level variables, as there was no data on whether those who were in accidents were watching football the prior night; further studies would be prudent before policies are changed based on such data; furthermore, taxi drivers are still presumed to work on weekends so it is unclear why moving game days would lead to sleeping in. What about policies by the taxi company to encourage drivers to watch recorded games rather than live or other incentives/disincentives for sleepy driving?

We think these are all valid and wonderful practical implications, which we discuss on pp. 13-14 in the general discussion. For example, we encourage future research to survey drivers who were in accidents and directly assess their football viewing habits and sleep hours. We also encourage more future research more generally to replicate and extend our findings.

Finally, with the new Taiwan data we are more confident that this effect is not specific to taxi drivers but to drivers within the GMT + 8 time zone more generally. We thus continue to encourage football associations to schedule high-profile games strategically on Friday or Saturday nights.

Conclusion about saving lives does not come from the presented data- if authors can look at traffic fatalities this would be justified but if not, then conclusions should be relevant to the data presented (accidents alone)

Because we do not have fatalities data, we no longer discuss lives saved. We only state that policy changes to football games can potentially reduce injuries related to traffic accidents.

Thank you for your constructive comments!
Reviewer 6

The Authors propose a novel hypothesis that sleep deprivation induced by watching high-profile football matches could induce more traffic accidents. They test this hypothesis using time-series and count model methods for a large taxi company in Singapore. They use as the independent variable the total market value of the salaries of European footballers playing on particular days. Results support the hypothesis and indicate a rise in traffic accidents during the day, following a night of a high-profile football match.

I commend the Authors for use of various methods to enhance methodological rigor, including the test which separates daytime from nighttime accidents. The hypothesis is novel, the proposed mechanism connecting football matches to traffic accidents is plausible,

the rigor of the study design and analysis is strong, and the Authors' conclusions do not overstate their findings. Below I list my comments, in order of importance.

We appreciate your kind words about our paper. We address your comments individually below.

1. Show ACF and PACF of residuals in traffic accident series before conducting cross-correlation function with the football match variable.

Granger-cause logic states that X cannot be said to cause Y unless it can predict Y better than the history of Y itself. In the cross-correlational analysis, you have not provided sufficient evidence that the history of Y has been modeled (and removed) BEFORE the effect of X on Y is examined. Please report the autocorrelation and partial autocorrelation functions (ACF and PACF) for the lags of your Y series after the alleged patterns have been identified and removed. You allude to a Dickey-fuller test for trend, but trend is but one of many forms of autocorrelation that should be removed from traffic accidents before the X and residualized Y series are cross-correlated. Removal can take many forms, including the addition of dummy variables, insertion of ARIMA error terms, etc. But the key here is that the history of Y is completely unavailable for explanation of Y on a particular day. Information on the ACF and PCF of the residuals of Y is standard reporting for time-series analysts.

Thank you for this point. You are right that the total accidents time series contains AR processes. In one sense, it is worthwhile modeling these processes because they are meaningful elements of the time series. For example, the negative AR(1) process may mean that people are less likely to get into traffic accidents following a day with high levels of traffic accidents, and it is conceptually unclear why we should remove this process when conducting bivariate analyses. It is much more obvious why we should remove trends, as these could lead to a spurious relationship between average market value and accident rate arising from both time series increasing linearly over time.

Nevertheless, we respect the importance of replicating the analyses under a variety of conditions. In our supplemental materials, we report an ARIMA model showing the AR and MA processes of the accident time series. We then take the residuals of this model and show that the ARIMA-residualized time series still is significantly associated with the average market value of football games in Singapore and Taiwan.

We summarize these analyses (and their meaning) on pp. 31-32 of the supplemental materials.

2. Policy implications unclear

Whereas I find the results to be compelling, the implications for policy are less clear. A full accounting of the costs/benefits of high profile football matches in GMT+8 surely would involve (I) potential health benefits of watching the game—for cardiovascular health, stimulating inspired individuals to exercise more the following day, opportunity to socialize and interact with peers and avoid social isolation); (2) economic benefits to Singapore for showing the matches—in terms of cable TV subscriptions, revenues for restaurants and bars

that remain open late, etc. Surely other avid football fans could come up with other potential benefits.

For these reasons, a lack of complete accounting of the net cost/benefit of these matches would make the policy implications of findings less clear. I would modify the discussion accordingly, and/or augment the cost/benefit analysis.

This is indeed an excellent point. We can only estimate the “costs” associated with watching football games but it is hard to estimate all of the “benefits” you mentioned. With that said, we are now much more careful in toning down our impact analysis, and clearly indicate that this economic impact analysis should be interpreted with caution because while there are economic costs there are also offsetting benefits such as the ones you mentioned. We then encourage future research (perhaps among economists) to better gauge the net cost/benefit of our findings (pp. 13-14). We hope this can assuage your concern on this particular issue.

3. Aggregation of start times assumes homogeneity in sleep response regardless of start time

Not all European matches start at 7pm local time. Aggregation of a high-profile match index, regardless of the mean (or most important) start time, assumes that a sleep deprivation response to a 2am (vs 4am) start time is the same. Would sleep research suggest similar responses in terms of total hours lost due to these different start times? This seems unlikely, given that 2am starts might interrupt REM sleep more than would 4am starts, and the ability to return to sleep after a 2am start may be more/less likely than would a 4am start. More information from the sleep literature would assist, so that the reader could determine whether it is reasonable to assume homogeneous sleep disruption responses to different start times of games.

The heart of your question appears to be whether a football game start time is more or less disruptive depending on whether it interrupts Rapid Eye Movement (REM) sleep or other stages of sleep. This is an interesting question. We do not have any data regarding the sleep stages of the taxi drivers. However, as you suggest, the sleep literature can be informative in thinking through the issue that you raise.

Sleep is a heterogenous state which can be broadly categorized into REM sleep and non-REM (NREM) sleep. A given individual will flip back and forth between REM and NREM sleep several times throughout a given night of sleep (1). Although REM and NREM sleep have relatively distinct methods of regulation as well as different specific functions (2,3), crucial recovery, maintenance, and growth activities occur during both REM and NREM sleep (4), as well as important memory consolidation functions (5,6). For example, synaptic homeostasis occurs primarily in NREM sleep (7), and emotional processing occurs primarily in REM sleep (8).

However, despite these distinctions, it is important to note that both REM and NREM sleep are important for healthy waking function (4), and thus interrupting and neglecting either stage hinders effective waking functioning. There is no basis in the literature for assuming that interrupting REM sleep is more harmful to driving safety than interrupting NREM sleep.

Nevertheless, this is a question worth pursuing in future research. Accordingly, on p. 13, we now note that future investigators should examine whether waking someone from REM sleep increases traffic accident hazard rates throughout the next day more than waking someone from NREM sleep.

1. J. Lu, D. Sherman, M. Devor, C. B. Saper. A putative flip-flop switch for control of REM sleep. Nature. 2006. 441, 589-594.
2. S. H. Lee, Y. Dan. Neuromodulation of brain states. Neuron. 2012. 76, 209-222.
3. P. Fort, C. L. Bassetti, P. H. Luppi. Alternating vigilance states: New insights regarding neuronal networks and mechanisms. European Journal of Neuroscience. 2009. 29, 1741-1753.
4. R. E. Brown, R. Basheer, J. T. McKenna, R. E. Strecker, R. W. McCarley. Control of sleep and wakefulness. Physiological Reviews. 2012. 92, 1087-1187.
5. S. Diekelmann, J. Born. The memory function of sleep. Nature Reviews Neuroscience. 2010. 11, 114-126.
6. B. Rasch, J. Born. About sleep's role in memory. Physiological Reviews. 2013. 93, 681-766.
7. G. Tononi, C. Cirelli. Sleep function and synaptic homeostasis. Sleep Medicine Reviews. 2006. 10, 49-62.
8. M. P. Walker, E. van der Helm. Overnight therapy? The role of sleep in emotional brain processing. Psychological Bulletin, 2009. 135, 731-748.

Thank you very much for your time and constructive feedback. This paper has improved dramatically as a result and we hope you agree.