Dear Editors of Seismica,

Thank you for taking the time to handle our manuscript during its revision process with Seismica. We are happy to hear constructive responses from the reviewers – we agree with their assessment that this work nicely contributes better defining how TLPs can manage hydraulic fracturing related seismicity in the UK. Generally, the critiques focused on clarifying our methods, placing caveats on interpretations, and digging a bit deeper into interpretations (sometimes with interesting new analysis). Above and beyond this, we have also added new analysis (Figure S7) that applies our 'bookending' approach to estimate risk tolerances from the PH-1 events, which agree with our PNR-2 estimates. As well, we have re-read the paper with fresh eyes to catch grammatical and typographic mistakes. We feel that our changes in the revised version are converging on a manuscript suitable for publication – we hope that you agree.

On the following pages are our itemized responses to the reviewer's comments. The original comments are in **bold-face***, our responses are in italics, and callouts to the original text are in "*blue-face.*"*

Thank you,
-Ryan, Brian, Benjamin, & Stefan.

Reviewer A:

This paper presents an interesting study into the potential impacts of shale gas induced seismicity in the UK, and how traffic light schemes for induced seismicity mitigation should be designed under a risk-informed approach.

In my view the restarting of a shale gas industry in the UK seems very unlikely (for political reasons, rather than any geological issues). However, the paper will still be of great interest since it presents a methodology that could easily be adapted to other technologies such as geothermal and/or CCS. I certainly believe that this paper will be of broad interest to the Seismica community.

The paper is well written and, on the whole (subject to a few caveats listed below) very clear. To the best of my knowledge, the work is scientifically correct. I am happy to recommend publication.

There are a few aspects where some additional details or clarity would be welcome, as described below:

We thank James for his kind words and time in reviewing our manuscript! We're happy to have his suggestions, which have improved the quality of our manuscript. Our line-by-line responses to each comment follows below.

Explanation of the underlying method. While I appreciate the authors are building on an already-published method here, a little more detail as to the underpinning theory is required. This is especially apparent in Section 2.1, and around the definition of Equation 1 in particular, where Rs needs to be explained properly (a 'count ratio' might not be clear to many who are not familiar with the topic), and u1 and u2 are apparently just "random variables" (not at all helpful as a description).

We have revised the sentences considerably in the first paragraph of Section 2.1 to address this and the following points.

My assumption is that the DM value (Equation 1), which is estimated stochastically via randomisation of u1 and u2, is added to the red light value in order to estimate the size of the event that occurs for a given choice of red light. But nowhere in the text is this actually stated explicitly. Presumably, in most Monte Carlo instances no trailing DM increase occurs (since you have 85% of

events during injection and 15% after), in which case is the modelled earthquake scenario given by the red light threshold itself?

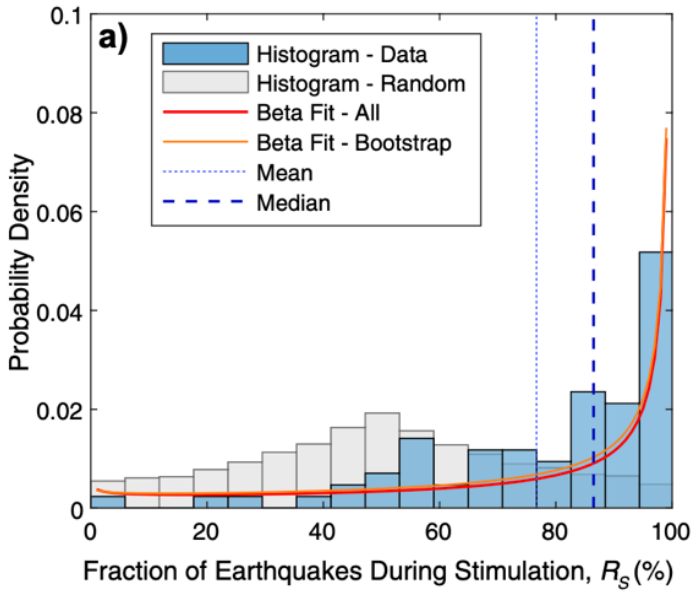
The reviewer is correct here regarding u_1 and u_2; these are uniform random variables. We have revised Section 2.1 sentences to be more explicit on how ΔM is stochastically drawn from u_1 and u_2.

Choice of Rs values. I don't think the choice of Rs value used is ever stated explicitly, and it is never explicitly justified except by reference to prior publications. Given the importance of this parameter, the choice of value used needs to be stated explicitly, and at least a brief justification given for how the value was chosen.

On a related note, given that most of the other parameters are chosen from underlying distributions using a Monte Carlo method, I don't understand why a similar approach isn't also used for Rs (i.e., allowing it to be selected stochastically from an underlying distribution).

On this point, we do state our choice of R_S values being drawn randomly on lines 166-169. However, we acknowledge that we could be clearer as to how Rs (and ΔM) was modelled in the paper. We have revised the sentences considerably in the first paragraph of Section 2.1 to address this point.

Briefly, we randomly draw an R_S value from a beta distribution. This beta distribution is based on the fit to R_S data for all the HF and EGS cases in the world [Schultz et al., 2022a]. For convenience, we've reproduced that plot below. From there, we draw a b-value from the distribution we described on lines 159. We can then use these two parameters in our Eqn 1 – where u_1 and u_2 represent to random numbers drawn from a uniform distribution. This is how we stochastically draw a ΔM value.



R_S distribution from Figure 2 in Schultz et al., [2022a].

My understanding of the method is that it assumes that the red light event occurs, and then calculates the probability of damage/nuisance/etc. under the condition that this event has occurred. This is a reasonable approach to take given the objectives of the study (i.e. defining regulatory limits), but the approach is really vulnerable and ripe for mischaracterisation. The various hazard curves and likelihoods of damage/nuisance could easily be misinterpreted as the likelihood of such things occurring if hydraulic fracturing takes place. Obviously, to estimate this, the computed hazards would need to be multiplied by the probability of a well experiencing seismicity up to the red light threshold in the first place.

Given the objective of the study and the method used its right that you don't do this, but a very clear statement is required to point out that these hazards and probabilities are conditional on having a red light event at the defined threshold, not the likelihood if any hydraulic fracturing takes place.

*The reviewer is correct here: we quantify the last possible stopping point before exceeding a tolerance to risk, assuming an earthquake would occur at that given location (while accounting for trailing seismicity possibilities). We agree with the reviewer that we could add a point clarifying this. We have done so on lines 264-267 with the statement "*We remind the reader that our approach quantifies the impacts that would happen following a red-light (including trailing seismicity) – it is unable to discern the likelihood of a red-light occurring, or the efficacy of mitigation procedures.*". While there may be*

some people who will mischaracterize the meaning of our results, we have been diligent about the limitations of our results.

For nuisance tolerance thresholds, I assume that the dashed lines in Figure 6 correspond to the values of tolerance used in Figure 7 (9571, 5478, 2719)? This needs to be explicitly stated in the text, presumably towards the end of Section 3.2.

Yes, these values are the nuisance tolerances that we derived from our 'bookending' method that were then used to define red-lights for the UK. We now mention this in the main text on lines 337-340. For reference, this information was also mentioned in the caption of Figure 7.

Also on Figure 6, please explain the x-axes. Is this the number of people (or properties) experiencing the given CDI level, with 'counts' (i.e., the y-axis) the number of times in the Monte Carlo modelling that this threshold is reached? This isn't currently clear at all.

Thanks to the reviewer for pointing this out. We have now clearly defined this in the figure caption. For reference, all the aggregate metrics (nuisance & damage) will be in terms of the total number of homes impacted.

In order to validate the approach, I do feel it is necessary to compare these models with the levels of damage/nuisance actually reported for these events in reality (e.g., the BGS did you feel it data). This need only be a qualitative comparison with observations, but I do feel that it at least needs to be mentioned and discussed.

This is an excellent point raised by the reviewer. To address it, we have added a new paragraph describing this comparison in the results Section 3.2 on lines 341-360. While this comparison is largely qualitative, our results are similar in comparison to the felt/damaging reports from the M_L 2.9 PNR-2 event.

We note that that some discrepancy between reported/modelled estimates is standard – as risk impacts often vary on a log-scale, with uncertainties of approximately an order of magnitude. As well, this understanding also informed our choice to use modelled estimates for tolerance constraints (instead of reported values), as any estimate biases between both the tolerance/red-light estimates will tend to cancel each other out. See also a similar comment from Reviewer B.

Finally on this point, somewhere in the caption for Figure 6 you need to use the word "modelled" (or similar), to make it absolutely clear to the casual reader that these are not observations of damage/nuisance from these events.

Agreed. We have added the word modelled to the caption of Figure 6.

Application to PNR-1z and PH cases. You show how the approach would have fared for PNR-2, but I'm equally interested to see the application to PH and PNR-1. Would your method have recommended stopping before the M 2.3 PH event? Would your method have recommended stopping PNR-1z early, even though ultimately this well did not cause unacceptable levels of seismicity?

We agree with the reviewer on this point that this is something that we could elaborate on further. This has been one of the major changes to the revised version of the paper, in fact. Changes made here are also based partly on suggestions from Reviewer B, who wanted to see some additional sensitivity analyses for how red-lights change as a function of R_S.

In the manuscript, we have now added a new analyses that provides a pseudo adaptive look at how red-lights would have adapted to case-specific information. Just making the simple assumption that the previous stage's R_S value is used in the place of the global average estimate. These new results can be seen in Figures 10, S11, & S12, as well as in Section 4.2 on lines 501-522. What we find is that the adaptive approach would stop the PNR-2 operation at stage six, before the final stage that caused the largest earthquake. This approach would not have stopped PNR-1z, which did not cause a concerning event. The data at PH-1 is insufficient to do an adaptive approach, so we are unable to make a claim here.

This is a crude and simple approach, done simply to demonstrate the sensitivity of red-lights to R_S values and the importance of doing so. There will certainly need to be future research efforts into understanding how to best do this updating process; thus, we have made appropriate caveats to the reader on this point.

Your conclusions in terms of optimum places to do further HF (i.e., along the east coast) are basically with respect to exposure and vulnerability to the potential hazard. The other part of this equation of course is the likelihood of experiencing said hazard, which is not considered in your study. I mention this because there is ongoing work to characterise variabilities in geomechanical conditions across the Bowland Play (stress conditions, fault densities, etc.), which could affect

estimates for where might be best to operate with respect to the likelihood of causing induced seismicity.

Hence, while the conclusions are fine, you should be clear that this is a conclusion with respect to the exposure to the hazard, not with respect to the likelihood of causing the hazard.
Some other minor details:

*Yes, the discussion section here (Section 4.3) was largely coming from the point of view of a prospective operator who has no information of which regions are susceptible to earthquakes or not. To be clear that quantifying the susceptibility could also aide in HF location siting, we have added the sentences "*We note that this approach focuses solely on the potential exposure to risks. Complimentary siting approaches that consider the likelihood of induced seismicity, depending on the geological susceptibility to earthquakes [Pawley et al., 2018; Hicks et al., 2021], could also aid in choosing safer HF locations.*" on lines 542-545.*

L25: "stepping stone to net zero" - whether or not shale was or could be a stepping stone to net zero in the UK is a contentious issue. I think the main objectives for pursuing shale gas in the UK were reasons of energy security and economic benefits. You'd be on safer ground couching it in those terms.
This is a fair point. We have removed this sentence from the abstract.

L34: East coast, not west coast.
Thanks for catching this oversight!

L64: Verdon and Rodriguez Pradilla (2023) (http://dx.doi.org/10.1016/j.tecto.2023.229898) would also be a relevant citation to the discussion on this line regarding variability in induced seismicity prevalence.
Agreed. We've added a citation to this paper.

L111: Correct reference is Verdon and Bommer, 2021.
Corrected (there and throughout the paper).

L158: Earthquake depths. In the UK, fracking is prohibited at depths shallower than 1,000 m. Hence, where formation tops are less than 1,000 m, you should use a minimum depth of 1,000 m.

Where the formation base is less than 1,000 m, you should not compute a hazard since fracking would not be allowed in the formation in question.

*We appreciate the reviewer pointing this out. We weren't aware of this rule at the time of writing. From prior sensitivity analysis [Schultz et al., 2021b], we know that depth is a very small factor in changing the red-light thresholds. Because of this, we have opted not to accommodate this change – it would require us to completely re-run and then revise the paper for what we expected to be a nearly negligible impact. That said, we now mention this point in the manuscript on lines 179-180. "*We do not account for UK legislation [UK Infrastructure Act, 2015] that prohibits HF operations shallower than 1000 m.*"*

Also in this section, you should clarify that while you are including the Weald in your study in case hydraulic fracturing is ever used there. However, at present there are only (and have ever only been) conventional activities in the Weald Basin, including the Kimmeridge. No hydraulic fracturing is used in these activities. Otherwise, this could be a cause for concern for any members of the public in the Weald who read your paper and see oilfield related activities going on nearby their homes.

To address this point, we have added this disclaimer alongside our rationale – which is to be comprehensive and give discussion about prospects on lines 182-183.

L273-276: This is the only place in the manuscript where there are a couple of poorly written sentences.

We have revised the sentences here for clarity. If the reviewer has more specific comments or suggestions, we'd be happy to consider them.

L390: If we are talking about a red light of ~ M1.5 and a 2-unit gap for yellow, then we'd be looking at some very low yellow light levels. You might want to comment on whether these levels would really be appropriate, if for no other reason than achieving such detection thresholds without very high quality monitoring networks might not be possible.

We agree that this could make monitoring requirements costly in certain situations. In some cases, this may even make the entire HF operation impracticably expensive. That said, the approach we've outlined is based on justified principles throughout. To compromise those principles would also compromise the ability to stay below these tolerances. Admittedly, this puts the operator/regulator between a rock and a

hard place. They must decide between if they're willing to change their tolerance to risk, willing to pay for the monitoring, or simply unwilling (or unable) to allow the HF operation.

To be frank, we feel that this is a strength of our approach. It makes it explicitly clear what kinds of risks are being taken, so that a candid conversation can be held with local communities – maybe they would be willing to accept larger risks, if they were informed ahead of time and came to an agreement.

Figure 5c - the colour appears to be flat across the image, despite a broad colour scale in the color bar. Is this intentional?

Yes, this is intentional. Relevant risk tolerances for LPR are in the range of 10^{-4}-10^{-6}. For example, the Netherlands has put into law that 10^{-5} is their safety limit policy. We truncated this plot at 10^{-10} as this is well below this value range. Essentially, we're trying to show here that we never even approach significant (local) fatality concerns with a red-light of M_L 2.5 – even when using conservative vulnerability functions.

Kind regards,
Dr. James Verdon
University of Bristol
Recommendation: Revisions Required
We thank James again for his time in reviewing our manuscript!

Reviewer B:

This work develops a set of red-light thresholds for hydraulic fracturing in the UK, based on risks of nuisance, damage and fatality. These red-light thresholds range between ML 1.2 and 2.5 and are controlled by risks from nuisance and damage. The manuscript deals with an important societal topic that would be of interest to readers of the journal. I therefore believe the work may be publishable, although I have some significant comments/concerns that I think should be addressed before this would be possible.

We thank the reviewer for their insights and time spent in reviewing our manuscript – which are appreciated and addressed below.

Major Comments

The biggest concern I have is that the magnitude thresholds obtained are dependent on some significant assumptions. While the authors acknowledge that the "workflow is adaptable and can incorporate new components or updates as needed", the main point of the paper is to quantify and provide red-light magnitude thresholds, so the validity of the assumptions is critical.

We agree with the reviewer that we could be clearer on outlining our assumptions and methods. We have taken care to thoroughly address each suggestion in the comments below. We would like to emphasize that the codes used to produce all these results are available online via GitHub (and Zenodo). Readers who are curious can find the exact details to how we produced our results – and even change values/assumptions to see how this impacts the results.

One of these assumptions is the trailing count ratio used, which is significantly different to those observed during actual HF operations in the UK (as acknowledged by the authors in Section 4.2) . To help the readers judge the significance of this assumption, I strongly suggest that the authors provide a sensitivity analysis on this number. How do the results change if much lower ratios are used? This is particularly important, given that lower values would result in lower magnitude thresholds.

We note that a sensitivity analysis has previously been performed on our method [Schultz et al., 2021b]. There we found that this is the most important metric to changing red-light thresholds.

More directly to this point, we have embedded a new sensitivity analysis (via a simple/pseudo adaptive TLP) to our Figure 10. There we show how the red-light could have changed, if the operator used the Rs value recorded from the previous stage (instead of from the global distribution). This is interesting on two fronts: it gives the reader an example of the sensitivity to Rs and places that information into the relevant context. We have added additional discussion around this new analysis in Section 4.2 on lines 501-522.

Outside of this additional analysis, we feel that a more rigorous treatment of how red-lights should change (or be adapted) as new real-time information is recorded is outside of the scope of this paper. We have mentioned this point in the discussion of Section 4.2. We hope that future readers may be interested to follow-up with answering this question!

The second significant assumption relates to the choice of fragility/vulnerability models. It is not immediately clear why the Groningen functions would be appropriate in a different context, and no justification is provided. It is hard for me to judge this, as I cannot find the Crowley and Pinho (2020) report online (the authors should provide an access link in the reference section). Note that an extensive description of appropriate fragility/vulnerability models for PNR is provided here – I think it is crucial to understand if your choice of fragility/vulnerability models aligns with that study. If not, then the validity of your assumptions needs to be explained.

To this point, the reviewer has a similar critique for the fragility and vulnerability functions used for our damage and fatality estimates, respectively. Here, we handle each of those two separately below.

For the vulnerability function, we note that we have also compared the Groningen vulnerability function against the USGS empirically derived version for natural seismicity in the UK (Figure S5). The Groningen vulnerability function is more conservative than the USGS-UK version. However, nuisance and damage concerns, from our analysis, are more relevant to the UK – even when using this conservative vulnerability function. Based on this, we don't feel it is relevant to dig further into the impacts of modifying/improving the vulnerability function, as it won't change the results or conclusions of our study. We have included a citation to Crowley et a., [2017] alongside the Crowley & Pinho [2020] report (which also has a URL link for its reference now).

For the fragility function, we now provide a brief justification for their use on lines 212-214 as well as a comparison between our modelled results and the reported impacts from the BGS 'Did You Feel It' reports (like a comment from Reviewer A). We justified the use of these functions since they are a small subset of

fragility functions that were intentionally calibrated for use with small-moderate sized (induced) earthquakes. In this sense, they are one of the best analogues available for our answering the intended goals of our problem. Comparison between modelled-actual (self) reports suggests that our fragility model is estimating DS1 well, with some underestimation for DS2 (see comment from Reviewer A). We note that we have intentionally used our modelled results to determine tolerances (rather than the reports), since this provides some forgiveness against methodological errors – if we are consistently under/over-estimating the impacts, then the tolerance and red-light estimates will both make this error (and they will cancel out).

Many further arbitrary assumptions are made without justification or adequate explanation – please refer to my next comment for more details.

In general, given the topical nature of this study and the fact that there is some non-negligible likelihood of it being used to inform policy-related documentation, I think it is incumbent on the authors to clearly highlight the substantial caveats of their results at every opportunity, particularly in the abstract and the non-technical summary.

We have taken the opportunity to review/revise our abstract and non-technical summary, considering this suggestion. If the reviewer has more specific thoughts or instance they'd like to point out, we would be happy to discuss them.

The second major concern I have is the general lack of detail provided on the calculations. For instance:
- No information is provided on the random variables used in equation 1, and I don't think the Rs values are ever defined explicitl

Like comments from Reviewer A, we have revised Section 2.1 to be clearer on definitions of terms and the methods applied to randomly draw ΔM values.

-The depth distribution is not specifically characterized (beyond a visual representation of depth perturbation), nor justified. It is not clear whether different depth distributions are used for each formation or one depth distribution is used for all earthquake sources. If is the latter, I would like the authors to comment on the appropriateness of this assumption.

On this point, we justified our choice of modal depth value and the distribution around that depth based on observation from prior HF induced seismicity around the world – HF-IS tends to be near the stimulation

interval, sometimes slightly above, but most often deeper. We explicitly mention this point on lines 185-186.

- I assume you use the Esposito and Iervolino spatial correlation model, but please make this more explicit.

The reviewer is correct here. We have now made this point more explicit.

- Why use a half Gaussian distribution for pre-building damage? Has this been used in other studies?

We use the half Gaussian to provide some level of uncertainty on the output of the fragility function – like the uncertainties included with the use of nuisance and vulnerability functions. We now explicitly mention this point on lines 230-231. This is to simulate some level of prior unnoticed DS1 level damage to a home. This has been done before, in a prior study of red-light thresholds from enhanced geothermal operations in the Netherlands [Schultz et al., 2022b].

- It is not clear what the "logic tree branches" and "coefficient covariances" of the vulnerability and nuisance functions are.

Our apologies for the lack of clarity here. These estimators have simple functional forms that are modified by their parameters. For example, the nuisance function is based on the logistic function. This function has two parameters: 1) that determine the location of the 50% chance point and 2) that determines the width of how quickly it transitions from 0-100%. We are perturbing these parameters to account for uncertainty in them. These uncertainties can be seen in Figure S4. We have revised the sentence on line 227 to be clearer.

- Line 214: "The population maps are adjusted to account for variation in population distribution." What does this mean and how was the adjustment carried out? Similarly (line 217), it is not clear how the population is adjusted to reflect temporal trends

What we are doing here is attempting to simulate variability in moving/commuting population throughout the day or uncertainties in our building inventory estimates. We use a Poisson-like distribution perturbation to change the household count value at the grid point. The revised sentence on lines 238-241

*now read as "*The population maps are perturbed to account for variation in population distribution and uncertainties in our household inventory; each grid point is perturbed by a Poisson-like distribution (Gaussian with a mean of the grid point's value and a standard deviation of the square root of the value), to account for these uncertainties*".*

- 214: Are the 400 and 40km distance limits epicentral, hypocentral, etc.? Are these limits derived from previous studies? It would be useful to know the shaking intensity/CDI values obtained at these distances, for some example magnitude and vs30 values

These are epicentral distances. We have noted this distinction on line 238 now. These cut-offs are chosen since moderate magnitude earthquakes (up to M4) will have negligible chance of damage or nuisance at those respective distances.

- Are each of the distributions sampled independently in the MC analysis or is there any covariance?

The variables are independently sampled. One exception being the fit parameters for the nuisance function, which have a covariance. This is now explicitly mentioned on lines 248-250.

- Line 239: It is not entirely clear that the "earthquake grid" is only defined around HF well locations

The earthquake grid is defined for all points within the shale play boundaries. There we take one earthquake grid point and simulate red-light events that are co-located with the operation. We clarify this on line 275 now.

This is a non-exhaustive list and I urge the authors to significantly expand on the details provided in Section 2 and the start of Section 3, to ensure the calculations are replicable.

Based on these comments, we have taken another look at Sections 2 & 3 to improve the details and clarity of our method. We feel that these sections are now clear to readers wanting to understand our work at a high-level. We have also provided significant details in the supplements to this paper, for readers who want to go into greater detail. Finally, we have also provided all the codes used to make the figures in our paper – so that future readers could also answer questions we didn't think to answer in our paper.

In my view, it is important to appropriately acknowledge previous work that is closely related to the topic of the study. Cremen and Werner (2020) also investigated nuisance risk from PNR events, and this should be recognized in my opinion. (Incidentally, the paper is included in the list of references but not cited anywhere in the text).

Thanks for pointing this out. We were aware of this paper and wanted to cite it, but it appears to have slipped through the cracks. We've now cited in on lines 422-423, where we justify the importance of considering nuisance risks.

The fact that nuisance and damage result in lower magnitude thresholds than fatality is self- evident; it is obvious that nuisance risk will occur at much lower magnitudes than fatalities. Thus, I am not sure why this is highlighted as a main finding of the study.

On this one point, we respectfully disagree with the reviewer. There is a bit of nuance that is important to recognize here. Generally speaking, this statement will depend on the tolerances for each type of risk metric. We consider a couple of hypothetical examples to make a point here. In the first example, if all the risk metrics had the same risk tolerances, then what the reviewer said is always going to be true – the risk metrics that require the least ground shaking will occur first, followed by the larger shaking ones (i.e., nuisance at CDI 2, CDI 3, CDI4, then damage at DS1, DS2, and finally LPR). However, if we have different values of tolerance, this may no longer be the case. For example, we could imagine a scenario where the tolerance for nuisance is 10,000 impacted homes and the tolerance for damage is 10 impacted homes – if we are in a remote region with disparately spaced towns of just 1,000 homes this nuisance tolerance will never be exceeded, but the damage one will. Again, we could imagine a similar scenario with only one person in a remote region – in this case LPR would be the first risk metric to be surpassed. This was a result drawn from prior work, in countries where the population density varied more dramatically – rural/remote municipalities had damage as their first red-light threshold [Schultz et al., 2021b; 2021c].

Minor Comments

"risk-exposure" is an odd term to include in a non-technical summary in my opinion. Since exposure is one component of risk, I'm not sure what it means. Furthermore, I don't believe a non-technical audience would necessarily understand it, so I suggest it be replaced with something like "exposure to earthquake shaking".

This has been reworded, as suggested.

The choice of intensity measures calculated in the hazard assessment for building collapse assessment directly depends on the intensity measure used in the collapse fragility functions. Please make this point clear in lines 180 to 184. Thus, the choice of intensity measure cannot be justified with reference to previous fatality risk studies, but must instead be justified on the basis of the fragility functions used to compute collapse probabilities.

We have added this clarification on line 200. We use PGV in our nuisance functions and fragility functions. An average of spectral accelerations over a range of values is used for LPR estimates.

Line 63: the word "susceptible" needs further elaboration - susceptibility to what?

Here we mean susceptible to causing an earthquake. On this point, we do feel that this is clear from the context of the sentence.

Line 275: "earthquakes significant HF events" -there appears to be a typo here

Thanks for catching this. We have corrected it.

Line 378: I wonder also if the discrepancy between the Horse Hill and HF thresholds is related to your use of absolute numbers. There is likely to be a much higher population around Horse Hill than around PNR for instance (particularly given the presence of Gatwick Airport near HH), so the discrepancies in absolute number thresholds may not be so significant if they were converted to relative values normalized by the total surrounding population number. I think a comment on this aspect would be a worthwhile addition to your discussion here.

On this point, the choice of absolute numbers for nuisance is intentional. We're interested in an aggregate metric, because the more people that feel the earthquake means there are more people who will be frightened and upset – potentially spurring a social change via their regulator. Using a local metric of nuisance would remove this.

Toward this point, we also just became aware of a recent study [Evensen et al., 2022] that surveyed/examined the attitudes of UK residents towards induced earthquakes. They found that a big factor

for influencing people's reaction to an earthquake was the anthropogenic source. This supports the results and conclusion of our study. We have added sentences on lines 449-453 to speak to this point.

6. Your recommendations for siting of HF do not consider future exposure changes (e.g., due to increased urbanisation). While I understand that a quantitative analysis accounting for exposure dynamics would not be necessary, I nevertheless believe that a comment on this limitation would be a worthwhile addition to the discussion.

This is an interesting point raised by the reviewer. With increased urbanization means more people and more homes in a smaller area – population/home density is going to increase with time. If you take our derived tolerances at face-value, this will mean that the red-light should correspondingly decrease with time. Generally, this comment speaks to interesting questions about the time-dependence of this issue and limitations of the tolerance model we've used.

We have added an additional paragraph, in Section 4.3, discussing this point (and its implications) on lines 566-576. This nicely serves as a launching point for future discussions around the management of induced seismicity.

Figure Comments:

Figure 4: Consider changing the lightest yellow colour used for one of the sites, as it blends into the yellow colour of the population colour bar.

Nice attention to detail here from the reviewer. We've inverted the colour of the yellow circle here.

Figure 5 (and all similar ones): I find the colours very difficult to discern on these figures. I would suggest having a much larger colour gradient for each scale, to enhance the readability. Furthermore, the caption mentions households impacted by DS1 and CDI 3. I assume the figure also captures households with higher DS and CDI values too?

To address this point, we have added contour lines to Figure 5, same as the iso-risk red-light figures. There are plots analogous to this for other CDI/DS values (Figures S9 & S10); however, we opted to place them in supplements to avoid bombarding the reader with extraneous information.

Figures 6 and S7: I assume that the x axis here (and indeed the risk metrics used throughout the study) refers to the number of households impacted by at least the CDI or DS value shown? Consider giving the x axis a more informative title.

Same as the comment from Reviewer A, we have now made it explicit that we are talking about the total number of impacted households.