We thank Tom Robinson for the helpful suggestions and comments, as well as taking the time to read and evaluate our manuscript. Please find outlined below our response to the suggestions made on the manuscript. We hope that we have added clarity where needed to the text. We have not transferred information from the appendix in the main text, as in response to the comments below additional information (not found within the appendix) was required to address the comments.

## Introduction:

A few more key references on landslides in S Alps would be useful – Korup, Davies, McSaveney etc all have plenty of articles relevant here that would be useful background.

We can include links and references to landslides in the Southern Alps in line 60:

*"In the Southern Alps of New Zealand, landslides are a common feature that play a significant role in driving erosion (e.g. Hovius et al., 1997, Korup et al., 2004) and present an increasing natural hazard and risk to people ad property (Allen et al., 2010, Cox et al., 2015; McSaveney, 2002)."*

It would also be good to see some more landslide QRA works referenced, at least breifly as there are certainly several around that would be useful to highlight

We can include a sentence to highlight the different uses and applications of landslide QRA in Line 33:

*"QRA are undertaken for a land use planning (e.g. Bell and Glade, 2004, Vega and Hidalgo,2017), infrastructure (e.g. Voumard et al., 2013. Macciotta et al., 2015), and for visitor destinations ( e.g. Corominas et al., 2019; Stock et al., 2014)."*

## Study Site

This needs an overview of at least the pre-covid number of visitors for reference. How many people on average visit per day?

We will include these numbers in the revised manuscript. For Fox Glacier Valley, the annual pre-COVID visitor numbers are approx. 400,000 per year with a maximum number of visitor's trips into the valley per day of approx. 3,500, while Franz Josef received approx. 700,000 per year with a maximum number of visitor's trips into the valley per day of approx. 6,000.

L75: Aseismic landslides needs a reference to support

We will add in the appropriate references.

## Method

Each representative earthquake event? Some more details on this would be good – is this just an Alpine Fault event or does this consider far-field sources too? What about potential seismic sources within the ranges (e.g. Cox et al 2012 – Tectonics)?

We will add more detail into the paragraph starting at Line 200, around the details of the national seismic hazard model and what it accounts for. In the NSHM, the active fault component defines the Alpine Fault local to Franz Josef as the AlpineF2K fault source. Within the NSHM the AlpineF2K source generates a $M_w$ 8.1 ± 0.2 earthquake with a single-event (strike-slip + dip-slip) displacement of c. 9.2 m with a mean recurrence interval of 341 years (Stirling et al., 2012). This is time independent variable and does not consider time elapsed since the last earthquake on the Alpine fault in 1717. Landgride et al., (2016) deaggregated the NSHM to see what other fault sources may contribute to the shaking hazard at Franz Josef. For a probability of roughly 10% in 250 years (or 2,500 years) the deaggregation indicates that the main contributor of seismic hazard is the $M_W$ 8.1 AlpineF2K source (i.e., the Alpine Fault). Additionally, the second largest seismic hazard over 2,500 years comes from moderate magnitude ($M_W$ 5-6) earthquakes that can occur <10 km from the townships. Although the Alpine Fault is the main seismic source in the area, the section of fault that could rupture might be located some distance away from the sites. For this reason, and to consider the contribution from the $M_W$ 5-6 earthquakes, we, therefore estimate the landslide severity for the four different bands of PGA as determined from the NHSM.

Compiled info on visitor / worker duration – could you expand this description a little here in the main body. This is crucial to understand some of the key variables to the risk equation. You've provided some nice details for the hazard part, so it would be good here to have some details on the exposure part. For instance, is this data pre-covid (something for the discussion). Is it averaged, or do you take demographics into account which may change exposure time (e.g. how did you determine an average walker vs a slow one?).

We will add more detail to Section 3.6. Our walking times for an average and slow walker was determined from DOC data and estimates. In Fox Glacier Valley, our average walker spent 1.5 hours walking to and from the glacier viewpoint and 0.2 hours driving to and from the car park (see Figure 1 c), while the slower walker spent 2 hours walking to and from the glacier viewpoint and 0.3 hours driving to and from the car park. In Franz Josef Glacier Valley, our average walker spent 2-hour walking to and from the glacier view point, and 0.3 hours driving to and from the car park (see Figure 1 d) while the slower walker spent 2.5 hours walking to and from the glacier viewpoint and 0.4 hours driving to and from the car park.

We also conducted our own field counts and checks on walking speed and approx. visitor numbers. This also revealed that not all visitors travelled the full length of the tracks and turned back at certain points. More detailed analysis and investigation of visitor behaviour and its impact on both exposure and vulnerability would be interesting to include but outside the scope of this study.

For the societal risk calculations (not discussed within this paper) we used pre-covid data of visitor numbers. Additionally, since 2019 and 2020 the access into Fox and Franz Josef valleys respectively, has been reduced having knock on effects to both individual visitor risk and societal risk.

Empirical estimates of vulnerability – largely agree, although the central estimate for 1000 m3 seems optimistic to me, even with evasive action

It is an average estimate for all landslide volumes below 1000 m3, as a 1000 m3 represents the bin boundary. As such, evasive action may contribute more to the vulnerability at lower volumes than at higher volumes within the bin, and as such for the central estimate is a representative value. However, we do account for the uncertainty by assuming a value of 1 in our upper estimate.

Seismic landslide inventories – are these 3 events likely to be representative though – rock types are similar enough as is the topography, but the climate is variable as is the earthquake history. Perhaps a point for the discussion, rather than expanding here in the methods but worthwhile all the same

We will include a point about this in the discussion under Section 5.1 regarding the importance of landslide inventories and uncertainties associated with the use of our 3 earthquake landslide inventories (see below).

*"For seismic landslides, the landslide inventories of the 2016 $M_W$ 7.8 Kaikoura, 1968 $M_W$ 7.1 Inangahua and 1929 $M_W$ 7.8 Murchison earthquakes (Massey et al., 2018b; Hancox et al., 2014, 2015), were used as proxies for Franz Josef and Fox Glacier Valleys given the lack of seismic landslide inventories for the West Coast. All three inventories were dominated by shallow debris avalanches, with such failure types potentially being the dominant type of seismic landslide type (Keefer, 2002). The schist rock mass of both glacier valleys is fractured with persistent faulting (Cox and Barrell, 2007) and therefore we assume that shallow debris avalanches are the dominant failure type. While all three inventories occurring in similar mountainous terrain to Franz Josef and Fox Glacier Valleys, climatic differences exist, with the impact of these climatic differences on the number and size of seismic landslides triggered unknown."*

L204 – Alpine Fault earthquake date needs a reference

Will add in the appropriate reference.

You've assumed landslides won't occur on slopes <30 deg – could you not use your compiled inventory to assess just how likely this is? Surely you have a slope frequency distribution you could use to inform this decision, or at least weight the probability component?

Our analysis of our landslide inventory shows that no landslides occurred on slopes less than <30 degrees. We will add this detail into Line 231. Slope angle is used within our susceptibility models to weight the probability component of the risk model.

Fig 2 – hard to read the legends and quoted power laws, particular in panel a

We will adjust to figures to make them more legible.

PGA input from NSHM – does this vary much over the valleys or is it pretty constant? Given the short valley lengths and distance from the Alpine Fault I would have thought there is little variation across the valleys, meaning it's the other factors that play the biggest role in determining landslide source?

The PGA input varies from 0.8 g to 1.1 g across both valleys. The other components of the earthquake induced landslide susceptibility model of Massey et al., 2018, such as distance to fault, slope angle, geology and local slope relief therefore also dictate landslide source probability.

Fig 3 – would be good to see the NSHM here to since that's a key input for the seismic landslides

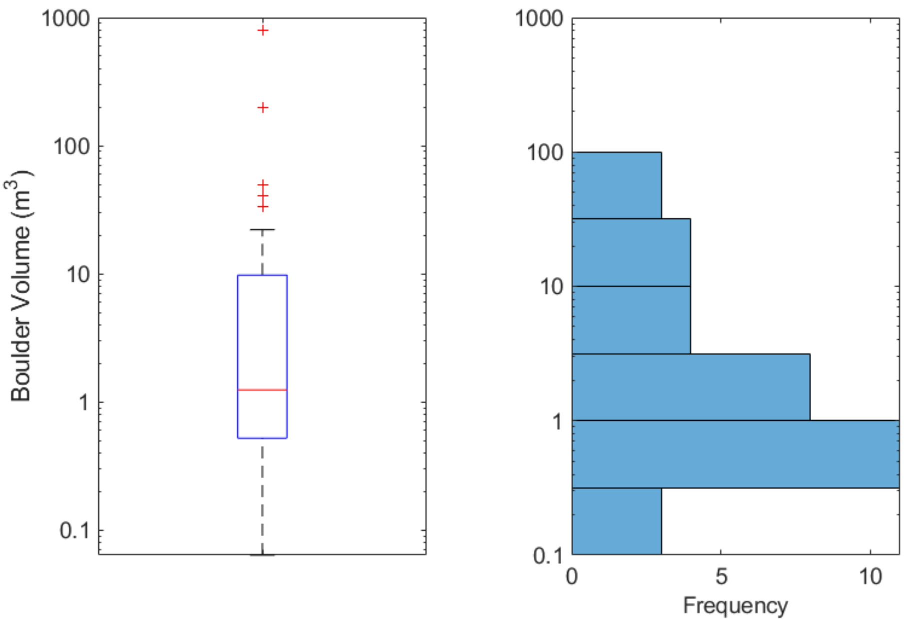
We can add in a box for the NSHM and the output earthquake induced landslide susceptibility model.

All slopes >45deg can generate rockfall – I don't necessarily disagree, but what is the justification for this?

We will add in the appropriate reference.

Field measurements show average boulder size is 1 m3 – again, would be good to see the distribution of this here in the main text somewhere to help support this - it would also be useful to see the range and skew of the data

We can provide more details in text and can include the figure below to display the data.




## Results

Fig 5 and 6 are very nice, but a more variable colour scheme would help, rather than graduated shades of blue which make it difficult to distinguish close classes

We will make the suggested change.

Fig 7 – its not immediately obvious that the y-axis scales differ here. At first glance I assumed the valleys were comparable. Could you either scale the axis, or make clear note in the caption

We will modify the axes to ensure that the y axis are the same scale.

## Discussion

If you think the order that you increase the variables influences the outcome, could you easily test this be changing the order and measuring the effect?

We have done this for one scenario, where we include increased earthquake annual frequency for Scenario 7 and Scenario 9. For example, the cumulative increase in risk associate with including a time dependent earthquake scenario in Scenario 7 is 330 % and 56 % for Franz Josef and Fox respectively, while for Scenario 9 it is 260% and 54%. In this example, we suggest that the differences are due to changes in the number of seismic landslides generated, spatial probability of impact and vulnerability. However, given that we are more concerned with the relative differences between the scenarios and therefore we think that the testing of the order of variables is out of scope and will not impact the relative differences between scenarios. We will update this point in the discussion to make this clearer.

The climate change discussion is a really interesting one and worthwhile here. However, the role of climate change on landslide rates in these areas is really complex and it's hard to confidently say what might happen – landslide frequency could drop while size increases for instance.

Agreed. We have tried to highlight this in our discussion but will expand upon this point and the associated uncertainties of increased landslide risk under climate change. In our sensitivity analysis we do not test changes in gradient of the -magnitude – frequency distributions to reflect increases in the frequency of larger landslides occurring relative to the frequency of smaller landslides. Such shifts in the magnitude – frequency distribution will impact the risk results and associated uncertainty.

One aspect missing from the discussion for me is the temporal variation in exposure. Firstly, covid may well have long term implications for visitor numbers that your values won't account for. Secondly, visitors in the valleys are no doubt much lower on rainy days when aseismic landslides are more likely than dry days, when aseismic landslides are less likely. The question is whether these changes cancel each other out. It's also not clear to me if you take diurnal variations into account or not – how many people are in the valley at night?

Will include a point regarding societal risk and the reduction in visitor numbers due to Covid. Additionally, within our societal risk calculations, detailed in Massey et al., 2018, we consider diurnal variations.

With regards to rainfall induced landsliding, we can include a theoretical reduction in aseismic landslide risk for visitor risk per trip that assumes that in heavy rainfall conditions when landslides are likely to be triggered the tracks are closed and the visitor is not present. Such a theoretical reduction was also included in the Massey et al., 2018 report. It is current DOC policy to close or partly close the tracks in each valley under heavy rainfall conditions or when heavy rain warnings are in place.

For this example (see figure below), we've set a theoretical reduction in aseismic landslide risk of 75%, assuming that 75% of our aseismic landslides are triggered under heavy rainfall conditions. However, as we cannot link our landslide inventory to rainfall events we can't provide a quantitative basis for this risk reduction number. Additionally, as mentioned earlier in our discussion section (Line 473), we may underestimate debris flow activity on the large debris fans and therefore are underestimating the rainfall induced debris flow risk on these fans. This again emphasises the need for a robust landslide inventory that captures all events. We will add in additional details and updated figure to Line 555 onwards.

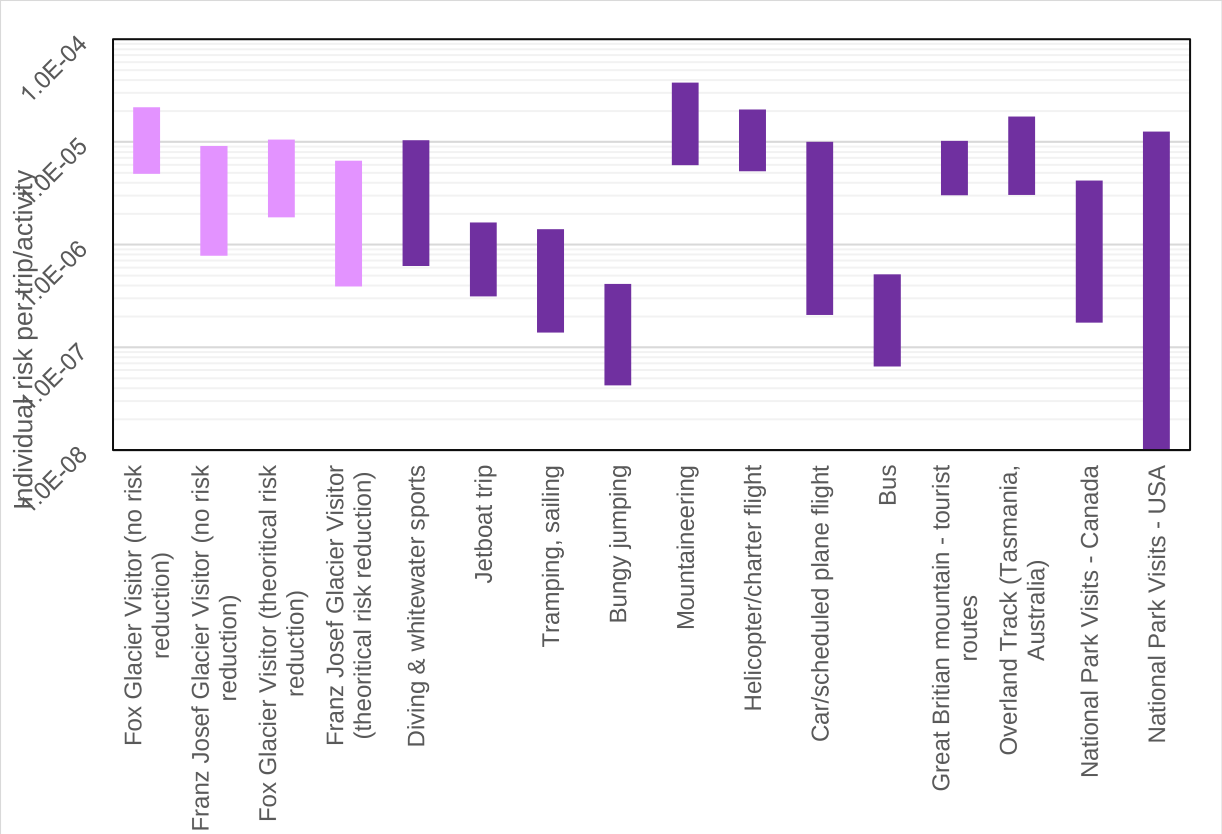



Fig 10 – excellent, very valuable. Could you maybe add the suggested 'acceptable' risk thresholds from the ChCh rockfall work for further added context?

Happy to reference the Port Hills acceptable risk thresholds in text, but as Figure 10 plots individual risk per trip rather than AIFR we will keep this separate to avoid confusion. The $10^{-4}$ AIFR corresponds to 3 x $10^{-7}$ risk per day