The paper presents a unique and interesting way of tackling the challenge of lack of information on hazard frequency and the implications for risk. The authors outline a methodological framework, using a Bayesian approach, to identify “windows of non-recognition” where debris flows return periods are enough to have not occurred in the recent historical record, but still may result in an intolerable risk to the population that have settled on the fan. This methodology has potential to be very useful for hazard and risk management. The paper is clearly written and easy to follow.

However, the manuscript as is requires major modifications, including more evidence to underpin some of the parameter assumptions as well as greater clarity about the purpose and key message of the manuscript. Further sensitivity analysis of the other key risk calculation parameters is needed to back up bullet point 4 of the conclusion. This sensitivity analysis may add to the strength of the argument for identifying the “window of non-recognition”. Throughout the manuscript there needs to be clearer links to the wider debris flow and risk literature. Re-structuring the introduction, discussion and conclusion may more clearly focus on the key message around the methodological framework and end purpose of the information.

There needs to be more links to the literature on risk evaluation and acceptable and tolerable risk thresholds. All the associated risk definitions need to have much tighter definitions. For example, it would be good to define what is included in unacceptable risk (i.e., intolerable and tolerable risk or just intolerable risk), as this was unclear to me throughout the paper. Unacceptable is term not often used or recommended given the two possible meanings stated above.

The same is needed for the calculations and definitions of individual and societal risk. Additionally, there is not a linear relationship between individual and societal risk. Societal risk calculation methods should follow those presented in Strouth and McDougall, 2020 and are scenario based. Is the method intended to calculate the Annual Probable Lives Lost (APPL) from the PH max event? And use APPL to back calculate PHmax. Given that you are evaluating the probability of only one debris flow scenario occurring (PHmax) this calculation route may be okay but requires further checking and a more in depth and robust explanation.

How does your stated risk threshold (which is based on individual risk) relate to societal/group risk, as these risk thresholds are often developed separately? Can this be explored in more depth? Linked to this is providing justification for considering both individual and societal risk, rather than just one or the other. How may this influence risk managers decision-making?

I was unconvinced by the assumption of the probability of spatial impact calculation. Is it possible to include within the sensitivity analysis an evaluation of this term, as it will likely have a big impact on the risk value. Can prior research such as Zubrycky et al 2021 provide distributions to evaluate within your Bayesian framework? This needs further links to the literature and explored more in the discussion. Is the Zubrycky et al., 2021 approach something that could be adopted in NZ?

Additionally, P S:H is a conditional probability of the hazard event occurring and therefore the equation 3 and 4 may need to be re-written to account for this conditional probability (e.g., 1 – (1-Prob) where prob is the individual hazard conditional probability).
Specific Comments

Line 35: Reference for “Debris flows as an unrecognised and underappreciated hazard...” Why is that the case?

Line 39: Remove the colloquial term “landslips”

Section 2.1: Move information about need for methodological framework to introduction.

Line 107: change from “is necessary” to “may be necessary”

Line 165: Missing references to paragraph

Line 189: Feels like this belongs in the discussion?

Table 1: Doesn’t match earlier description in text. Would be good to provide a separate overview table of published debris flow case-studies and associated ARI in NZ

Line 330: Missing references for this section. Pollock and Wartman 2020 – Human vulnerability to landslides might be useful here.

Line 335: Is this because we can’t always capture dynamic risk parameters (e.g., exposure and evacuation with heavy rainfall)? It would be good to highlight the need for dynamic risk models.

Literature mentioned in comments:

Strouth and McDougall, 2022 – Individual risk evaluation for landslides: key details

Zubrycky et al., 2021 – Exploring new methods to analyse spatial impact distributions on debris-flow fans using data from south-western British Columbia

Sim et al., 2022 – A review of landslide acceptable and tolerable risk

Pollock and Wartman, 2020 – Human vulnerability to landslides

S de Vilder