

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/nhess-2021-324-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2021-324

Anonymous Referee #2

Referee comment on "Travel and terrain advice statements in public avalanche bulletins: a quantitative analysis of who uses this information, what makes it useful, and how it can be improved for users" by Kathryn C. Fisher et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2021-324-RC2>, 2021

Thank you for inviting me to review this paper. I found it both very interesting and well written.

The aim of this study is to evaluate the use and effectiveness of the travel and terrain advice (TTA) statements in daily avalanche bulletins. More specifically, the authors set out to identify which user groups that pay attention to the TTA, how useful different user groups find the TTA, and if the usefulness can be increased by minor changes in the phrasing of the message.

The study addresses an important question for avalanche warning services (AWS) and other natural hazards warning services, i.e., how can we make the public warnings more useful and understandable for the targeted population? The authors use a relatively large and heterogenous sample of backcountry recreationists (N = 3100). This increases the chances that the results are generalizable to the general population of users of the avalanche bulletin. The authors employ a multilevel ordinal regression model (mixed effects model). Since the outcome variables are ordinal, and since the data is grouped on both the individual and TTA statement level, this seems like a reasonable approach. However, as described in the comments below, I also think that the estimation procedure creates challenges that are not completely addressed by the authors. Finally, I really like that the authors use interaction effects, as this allows them to analyze differences in effects in different user groups.

In conclusion, I find that this paper presents research that represent a substantial contribution to the literature on risk communication, and therefore contributes to our understanding of natural hazards and their consequences (Scientific Significance: level 1). The method used is valid. However, I think that some robustness checks are needed and that the discussion of the results would benefit from some restructuring (Scientific Quality: level 2-3). The presentation quality is excellent (level 1).

Comment 1

In spite of the relatively large sample size, the distribution of responses across both attention to the TTA and avalanche bulletin user types is very skewed. Most importantly, only nineteen participants stated that they pay no attention to the TTA. Ordinal models are sensitive to the number of observations in each cell. The low number of observations in the first cell of attention to TTA may create problems. This problem is amplified by the fact that the distribution of participants in the different avalanche bulletin user types is also skewed. There are only 15 participants, who self-categorize as type "A". Of these, two pay no attention to the TTA.

The low number of type A bulletin users may cause problems in all models in the paper. You use type A as a reference group in your analyses. In other words, the other user types are compared to type A users and not to other types. You in general find that avalanche bulletin user type is an important predictor in all models. However, based on the results presented in the tables, it seems like the coefficients on type "B"- "F" are very similar, i.e., the main effect appears to be between "higher than A" and "A". Since there are only 15 participants of type "A", these individuals are given a disproportional weight.

Suggestions

1) I would like to suggest that you check how well the models fit the data, e.g., by comparing the predicted probabilities in each cell (level of attention), with the actual distribution in the data. This is especially important for the TTA attention model. If you find that the model over-predicts observations in the lowest cell (no attention), I suggest combining this category with the next lowest category (little attention), and re-running your analysis on this variable.

2) To evaluate if the difference between type "A" users and higher level users is driving the effect, I suggest two approaches: either combining type "A" with type "B", or run the model on the subsample of participants of type "B" or higher. This comment holds for all models in the paper. The robustness tests can be included in an online appendix if the results do not differ in a significant way.

Comment 2

The correlation between avalanche training and avalanche bulletin user type is relatively high. For example, there are no participants of type "A" with advanced or professional

training, and no participants of type "F" without professional training. This may bias the results. The correlation may cause problems with inflated variance, and you will have some cells with zero observations.

Suggestions

I recommend comparing three different models: One where you include both type and training, one with just training and one with just type, and compare the models in terms of goodness of fit.

Comment 3

You provide a nice description of your estimation approach, i.e., using a mixed effects ordinal model. However, for someone who is not used to these models, the random effects in the models are not completely clear.

Suggestion

I think that it would help the reader if you write out the specification of the different models in equation form. This would make it easier to understand the random effects in the models.

Comment 4

As I understand it, you have two types of paired statements: 1) statements without extra explanation, with more or less jargon, 2) statements without jargon, with and without extra explanation. Together, you have four types of statements. In the regression in table 3, you use "more jargon" as the reference level. I have no problem interpreting the effect of reducing jargon, but I do struggle with interpreting the effect of "no added explanation" and "added explanation". How should I interpret these (insignificant) effects? If I have understood things correctly, then the "no added explanation" statements and "more jargon" statements represent distinct categories because the "no added explanation" statements do not contain jargon. However, the "no added explanation" is also different from "less jargon" because the original "no added explanation" statements did not include jargon. Have I understood this correctly? But then, how can you tell what it is that you are testing for here? It is quite possible that the problem is my lack of understanding.

However, I would like to see a more elaborate description on how to interpret these results.

Suggestion

I think that it would be substantially easier to interpret the results if you split the sample between the two treatment types (jargon and explanation), i.e., that you treat your data as emanating from two different experiments.

Comment 5 (minor)

Have I understood it correctly that you treat attention to the TTA as a continuous variable in the regression in table 3, while you treat it as an ordinal variable in table 2? What is your motivation for treating the variable in different ways in different models?

Comment 6

I really like that you illustrate the marginal effects in graphs. I also appreciate that you specify for which values you have estimated the marginal probabilities, and that you highlight that the reader should focus on significance levels and not predicted values. However, since the specifications used to estimate the marginal predictions are only given in footnotes, and since you talk about a broader group in the main text (e.g., people with introductory avalanche education), some readers may get the impression that your predictions are more general than they are.

Suggestion

Write out the specification in the text or in the figure caption.

Since there are some variations in the values chosen to estimate the marginal effects, I also think that it would be beneficial if you include a brief description of why you have chosen these values. Don't get me wrong, I find the chosen values reasonable as they, in some sense, represent a representative participant.

General comment

The main aim of the study is to test if it is possible to improve the understanding and usefulness of the TTA. However, the manuscript is structured in a way that draws the attention to the effect of avalanche bulletin user type on the outcome variables. Given both the stated purpose of the paper, and the issues with this variable (the low number of type "A" users), I think that the structure is a bit unfortunate.

Suggestion

Restructure the text so that each result section starts with a discussion of the effects of the main explanatory variables (e.g., jargon and added explanation).