

Responses to comments on egusphere-2022-244

"Performance of the flood warning system in Germany in July 2021 – insights from affected residents" by Annegret H. Thieken et al., EGU sphere, 2022

In the text below we provide a point-by-point response to the comments made by the referees. For clarity, our responses are written in Italics.

Referee 1: Martijn Kuller (Referee)

<https://doi.org/10.5194/egusphere-2022-244-RC1>

Referee comments

This paper presents the results of a survey held among the residents of a recent severe flood event in Germany. It is a valuable piece of work both for the academic community interested in risk communication and for authorities to draw lessons from and improve practices in the future. While the overall quality of the work sufficient to warrant publication, there are a few issues that need to be addressed. Most importantly, the structure of the paper could benefit from clearer research questions, hypotheses and a methods and results section following this structure.

Response 1: Thank you for valuing our work. We will consider changes in the structure of the paper and will phrase our research questions clearer in the revised version as outlined below. Initially, we did not follow a distinct hypothesis-based research design, but were more explorative. However, since we had some implicit hypotheses, we will phrase them explicitly in the revised paper. We will also carefully re-consider the structure of the individual sections of the article.

Overall the language is good and the paper reads well, but some phrases are rather “spoken” language than written language style (e.g., line 39: “save people’s lives”. I would rewrite to: “prevent fatalities” or “prevent loss of life”. Second example in line 79: “...how the targeted population...”. I would change to: “...warning perception by the target population...”). The article would benefit from review by a native English writer. Below I provide further details on this, as well as other comments by section, and then detailed comments in a table.

Response 2: The revised manuscript will be proofread by a native speaker before resubmission. The sentences pointed out by Reviewer 1 will be rephrased accordingly.

Introduction

This section has a clear structure and explains the background, the need for the research and the research aim clearly. However, detailed research questions and hypotheses are missing, which are needed to justify the methods used (types of data (indicators) gathered, methods of data collection etc.).

Response 3: In the revised version we will explain our research approach and underlying hypotheses clearer. Accordingly, we will extend the last paragraph of the introduction. A draft reads as follows:

“Therefore, an evaluation of a FFWRS should include how the addressed population perceived the warnings and whether they were able to respond adequately (Penning-Rowsell and Green, 2000). As part of a broader post-event investigation, this paper aims to analyse how the

warning system in July 2021 performed – also in comparison to other flood events in Germany. The evaluation of the performance of the warning system mainly focusses on two aspects: 1) How many people did receive a warning before they were in danger? 2) How well did affected people understand the warning and behaved accordingly? 3) How did the FFWRs perform in comparison to other recent pluvial and fluvial floods in Germany?

As highlighted by Thielen et al. (2022) for the river flood of June 2013 in comparison to the pluvial/flash floods of May/June 2016, the performance of Germany's FFWRs depends on the flood type. For pluvial and flash floods in 2016, there was a higher share of affected people who was not warned, warning times were shorter and the factual knowledge on adequate behaviour was poorly developed among affected residents. Given the severe impacts in 2021, we hypothesize that the performance of Germany's FFWRs in July 2021 was even worse than during recent pluvial and flash floods.

Since elderly people were considerably overrepresented among the flood fatalities of 2021, we expect that age is an important explanatory variable for the performance of the warning system next to the severity of the event.

Following an explorative approach, we also discuss how to further improve the FFWRs based on the outcomes of the analyses and the views and wishes of the population affected in July 2021 as mentioned in the survey."

The results section is already structured with regard to these questions/hypotheses. The methods section will be adapted.

Data and Methods

It is stated that the survey was distributed via facebook, and a press release was sent through the newspaper. However, it seems like the survey could only be accessed through facebook. Why did the authors choose this method, and how did they make sure this didn't lead to a biased sample, as facebook is only used by a specific demographic, leaving out other specific demographics that might be important. Beyond age, general techsavvyness and online presence play a role in the accessibility of the survey to people using facebook. If the authors somehow tried to correct for this bias, please provide some information about how this was done. If not, please discuss how the method might have influenced the outcome and what this means for the conclusions.

Response 4: This is a misunderstanding. The online questionnaire was accessible through a link outside of Facebook. Facebook was just used to advertise our survey. To avoid bias we also used traditional media like newspapers to draw attention to our survey and directly informed all mayors in the affected regions about our survey together with a plea to mention the survey in local newsletters. We saw a response to all advertising activities in the number of respondents to our survey. The age range presented in the paper illustrates that we managed to address elderly people, too, who are, however, underrepresented in the sample, although elderly people are increasingly online and use internet sources

We would like to highlight that we decided for an online survey in order to allow respondents a voluntary participation in the survey. This was important to us given the rather short time-span after the flood event, when many people were still in the immediate recovery process.

Also, it is explained what the questions in the survey were regarding, however I am missing some explanation as to why these questions were posed (what were the exact research questions the authors want to answer with the set of questions they selected, including the

demographical questions), and what were they expecting to find? Part of this information could already be provided in the introduction. Furthermore, is the survey accessible to the reader somehow? It would be good to add them to supplementary materials for reference and future use by the readers.

Response 5: With the revised introduction the content of the questionnaire will be more understandable. We will better explain the links between research questions and survey questions in the revised version. A translated version of the questionnaire can be provided as supplement if wished.

The Likert scales used in the survey are analysed statistically, using the answers as numerical values. Although this is common practice, it is strictly speaking not valid, as these numbers are only category indicators, and not actual numerical values of the answer, and can thus not be used for statistical analysis (drawing mean, sd etc).

Response 6: We use scales that are only verbalized at the end points as shown in Table A1. According to the (German) textbook of Porst (2014: Der Fragebogen, 4th edition, chapter 6), data from such a scale can be regarded as interval-scaled data (being ordered with equal distances between the values, but without having an absolute zero point). Therefore, it is valid to treat them as numerical data as it is also common practice, as also stated by the reviewer.

Results and discussion

If the factors investigated and presented in Table 2 explain so little in the receipt of an official warning, what then is (or could be) the most important predictor for this? All the results are self-reported data from the survey. I wonder how this self-reported data compares to behaviour and if this could be discussed in this section (e.g., how do the numbers of “knowing what to do” compare to the observed behaviour of people on the day?). Would there be any such comparison possible here in the discussion?

Response 7: Just for clarity: Table 2 shows the results of the regression on the receipt of the warning, while Table 3 shows the regression on factual knowledge (knowing what to do). For empirical data in social sciences, the latter has a good R^2 of 33 %.

We agree that it would be great to have some independent data to compare our results with. Unfortunately, there is only anecdotal evidence. For example, in the most affected district of Ahrweiler in Rhineland-Palatinate around 18 % of the population have subscribed to KATWARN, the commonly used warning app in this district. A more detailed analysis of our data for this district reveals that around 20 % of the people from this district in our survey report that they had been warned by this app. In addition, the time of having received the first warning as reported in the survey also matches to the officially released warning message.

Furthermore, during an interview with a representative of the regional broadcast we discussed the percentage of people, who were not warned. That number was then compared to data on the general coverage of TV and radio. According to the media expert a higher percentage of people would only have been reached if the warnings and the upcoming event had been addressed in the programme for a few days before the event happened and by using easily interpretable stories and images. The fact that warnings of slow onset fluvial floods like the one in 2013 are much more successful (as is shown in Thieken et al. 2022 and in this paper, too) was explained by the better coverage in the media for several days and first stories of affected places and people in the upstream areas.

Hard facts and data are, however, absent, but the aspects mentioned above can be briefly mentioned in the revised paper. In general, behavioural observations during the actual event would have been neither possible nor ethical.

The results presented in Figure 5 show very little variation. This is a common problem with asking respondents what they want, choosing from a list of options: they are going to want it all. Such results are not very helpful when resources are scarce and trade-offs have to be made (in this case, we want to keep a warning clear and concise). A better way to measure such preferences could therefore be asking respondents to rank or to distribute a limited number of points among the options.

Response 8: Thank you for this remark and the suggestion which we will consider in a next survey on this topic. Still, we think that our data are of use, e.g., with regard to the expected amount of rainfall. Here a clear mismatch between the frequency of this information in the received warning messages and the wishes for the future becomes apparent.

A section on further research is missing. I already made some recommendations in this review, e.g., comparing self-reported information to actual behaviour (if this is not going to be implemented by authors).

Response 9: We will add a dedicated paragraph with ideas on future research in the concluding section.

Conclusions

Because the research questions weren't completely clear (see other comments), the conclusion doesn't naturally follow from the rest of the paper. When this issue is resolved, the conclusion might be written in a more concise manner, going back to these research questions.

There are many references to the main text as well as other papers in the conclusion. I recommend rewriting the conclusion according to my above instructions, and avoid any references, as a conclusion should stand on its own, not repeat the results, but rather state the significance and meaning of results. As it stands, the conclusion mostly summarises the main text.

Response 10: Thank you for this suggestion. However, we partly disagree. A conclusion should or could start with repeating the research aim and the main findings, then proceed with implications (while accounting for the inherent uncertainties) and finally addressing future research topics. Following this structure, references to other works or to tables/figures in the paper might help the reader to connect things. According to the reviewer's preferences we will, however, reduce the number of references in the concluding section.

Furthermore, I think the paper could benefit greatly from a separate section (within conclusion or at the end of the discussion) that summarises the most important recommendations from this research. This research is very practical and applied, and has great potential to aid authorities around the world.

Response 11: In our view these recommendations are part of the conclusions as outlined above. We will pay attention to better address these in the revised version.

Please find detailed comments in the table below:

Line 39-40: One or two sentences discussing why the performance might have been so good for these events could benefit this section here.

Response 12: We will add some figures on the performance of FFWRS in the revised version.

Line 153-158: This belongs more in the methods section, see previous comments.

Response 13: These sentences will be shifted to the method section.

Line 159: Results are presented for which no methods have been described (logistic regression is mentioned). This should be added to the methods section (also see previous comments about methods).

Response 14: Logistic regression is a standard method. Therefore, we didn't mention it explicitly in the method section, but we will briefly do so in the revised version together with the paragraph from lines 153-158.

Line 160: Please explain how to interpret odds ratios.

Response 15: We apply an ordered logistic regression. The interpretation of the ordered logit coefficient is that for a one unit increase in the predictor, the response variable level is expected to change by its respective regression coefficient in the ordered log-odds scale, while the other variables in the model are held constant. As an intuitive interpretation is difficult due to the log-scale, we consider to provide odds ratios in the revised version of the manuscript as a measure of the effect size, which are easier to interpret. An odds ratio above 1 then indicates that, as the explanatory variable increases, the odds (or likelihood) of the dependent variable occurring also increases. Conversely, an odds ratio below 1 indicates that, as the explanatory variables increases, the likelihood of the dependent variable occurring decreases.

Table 2 and 3: From the table it isn't clear how gender impacts the prediction on knowing what to do. Is it male or female that increases the knowing what to do (explain in the table).

Response 16: The coding is explained in Table A1, but we will repeat the most important information in the headings of Tables 2 and 3.

Line 184: Is there data available to make a more precise comparison between the warnings issued by the different authorities and the warnings received by the population as reported through the survey? This could shed light on the effectiveness of the dissemination by the authorities. (a hint to this information is given in the text in lines 183-184).

Response 17: Please have a look at response 7 above. We haven't done this regional analysis for the whole region, which is quite large, but we could highlight one example to illustrate the finding.

Line 189: The authors state that they find the trust in the credibility high. I would argue that 9% of people not believing a warning is quite a high number in an emergency situation, and less

than half of the people stating they find the warning highly credible is to my perception quite low. From a government-issued warning, I would strive for credibility numbers close to 100%.

Response 18: We agree and will rephrase the sentence. There's anecdotal evidence that a lot of people distrusted the predicted severity, not the warning itself.

Line 198: This sentence is confusing, please re-write.

Response 19: The sentence now reads: "In addition, some respondents complained that too many warnings on Covid-19 were disseminated via the most popular warning app NINA, which was tiring and lowered attention to warning messages in general. Above all, in the week prior to the severe flood event there were already warnings for heavy rain in parts of the affected region, but had no serious (flood) impacts happened.

Line 210: This is more a concluding remark.

Response 20: This aspect will be included in the concluding section.

Fig. 5 and 6: Meaning of the numerical scale points need to be explained in the caption. Furthermore, as indicated before, it is strictly speaking not correct analysing and presenting this data in this way.

Response 21: Our apologies for this. The meanings of the scales are explained in Table A1, but will be repeated here. See our comment (response 6) on the validity above.

Line 298: It is not common to refer to tables and figures in the conclusion if not strictly necessary.

Response 22: see response 10

Thank you for the detailed comments to our paper.