Deliverable D5.3

Designing & implementing the seismic portion of dynamic risk communication for long-term risks, variable short-term risks, early warnings

Deliverable information	
Work package	WP5: Society: Data Gathering and Information Sharing with the Public and Policy-makers
Lead	University of Cambridge
Authors	Sarah Dryhurst, Giulia Luoni, Irina Dallo, Alexandra L.J. Freeman, Michèle Marti
Reviewers	Rémy Bossu, EMSC
Approval	Management Board
Status	Final
Dissemination level	Public
Will the data supporting this document be made open access? (Y/N)	N/A
If No Open Access, provide reasons	No original data is contained within this review
Delivery deadline	31.08.21
Submission date	31.08.21
Intranet path	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821115

Table of Contents

Introduction

Seismic information and knowledge is only of use if it helps inform the planning, preparation, and response to earthquakes. This requires careful, precise, honest and unambiguous communication - sometimes to people who have little or no grounding in geological principles.

Development of such communications requires careful co-design with the different audiences: an iterative process to understand their information requirements, prior knowledge and misperceptions, and then test the comprehension and usefulness of potential designs at every stage.

Such a process has been used, for example, to design “aftershock” forecast communications in New Zealand (e.g. (Becker et al., 2020, 2019; Wein, Potter, Johal, Doyle, & Becker, 2016) and the US (e.g. (McBride et al., 2018; Michael et al., 2019), and multi-hazard communications in Switzerland (Dallo & Marti, 2021; Dallo, Stauffacher, & Marti, 2020).

Here we describe the first stages of the iterative development of seismic communication materials to inform hazard planning and preparation, operational earthquake forecasting, earthquake early warning and rapid loss assessment, in three European countries (Italy, Switzerland and Iceland). These processes have attempted to tackle a number of the key communication problems inherent in seismic information, such as the basic probabilistic nature and high uncertainties in many of the quantitative measures, the small probabilities, and the trade-offs between useful degrees of precisions, comprehensibility of the numbers, uncertainties, and information overload.

Methodology

We applied a user-centred, mixed-methods approach to co-design and evaluate earthquake communications with target audiences in three European countries: Italy, Switzerland and Iceland. Figure 1 summarises the work described in this document. In parallel, we iteratively exchanged ideas and designs with scientists from different fields and improved our designs based on their ideas and feedback as well as that of the formal study participants, and conducted literature reviews across a range of domains (see RISE Deliverable 5.1)

The research will continue to follow the approach of mixing qualitative interviews/focus groups and quantitative surveys to further refine and evaluate communication products.

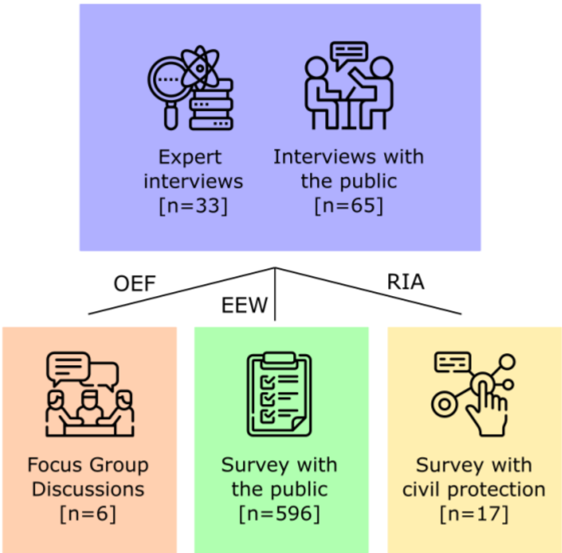


Figure 1: Studies conducted to analyse and explore the communication of dynamic earthquake information. The findings of these studies fed into the design of the products described in this document.

Long-term Hazard Maps

Long-term hazard maps are maps that display the likelihood of particular hazard events occurring over a long timeframe, typically tens to hundreds of years. Maps are popular as a method of communicating earthquake information (Gaspar-Escribano & Iturrioz, 2011), and several countries produce such maps for communicating earthquake hazard. Indeed earthquake hazard maps are the most common way of visualising seismic hazard in the long term (Marti, Stauffacher, & Wiemer, 2019). Although such hazard maps are typically designed for expert users (Perry et al., 2016), seismic hazard maps have many other, non-expert users including the general public. Hazard maps are usually communicated in an unaltered form to non-expert users however (Marti et al., 2019; Thompson, Lindsay, & Gaillard, 2015), and there is evidence from other disciplines that hazard maps are poorly understood by the lay public (e.g. (Hagemeier-Klose & Wagner, 2009; Perry et al., 2016). A notable exception are the seismic hazard maps used by the Swiss Seismological Service (SED), which were evaluated by (Marti et al., 2019) (see later).

Current products

United States Geological Service, USA

The United States Geological Service (USGS), releases a map within the USA that displays the likelihood of peak ground acceleration being exceeded within 50 years (see Figure 2).

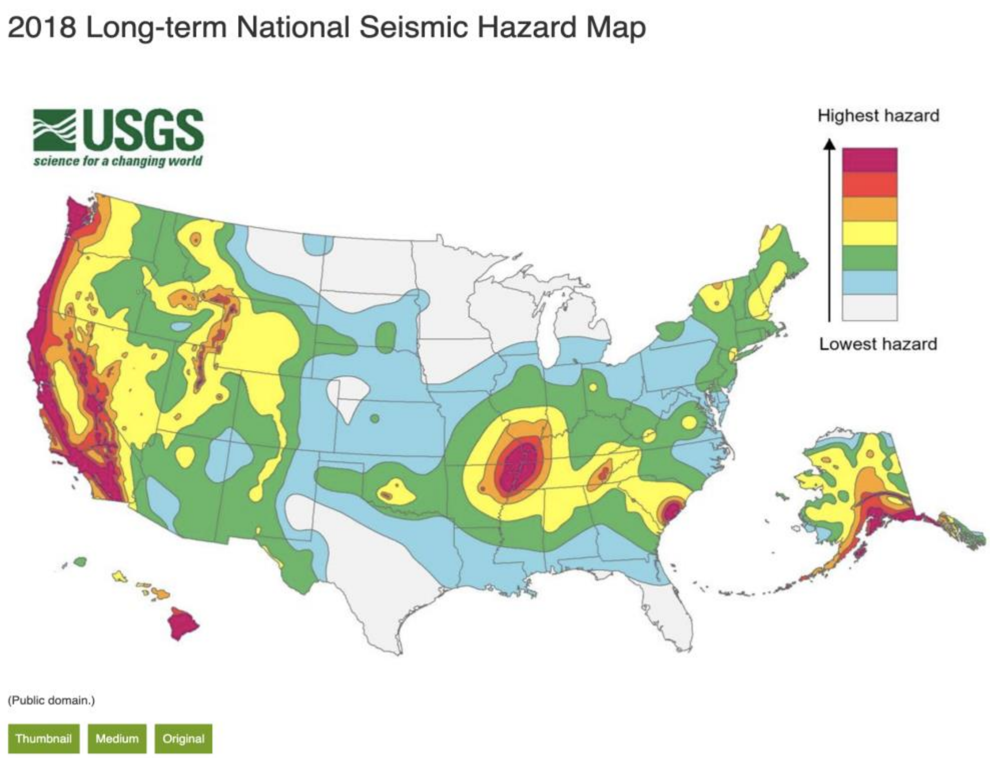


Figure 2: Earthquake hazard map showing peak ground accelerations having a 2 percent probability of being exceeded in 50 years, for a firm rock site. The map is based on the most recent USGS models for the conterminous U.S. (2018), Hawaii (1998), and Alaska (2007). https://www.usgs.gov/media/images/2018-long-term-national-seismic-hazard-map

In addition to this, the USGS also reports the expected number of earthquakes exceeding intensity (Mercalli scale) VI or higher within the next 10,000 years (see Figure 3).

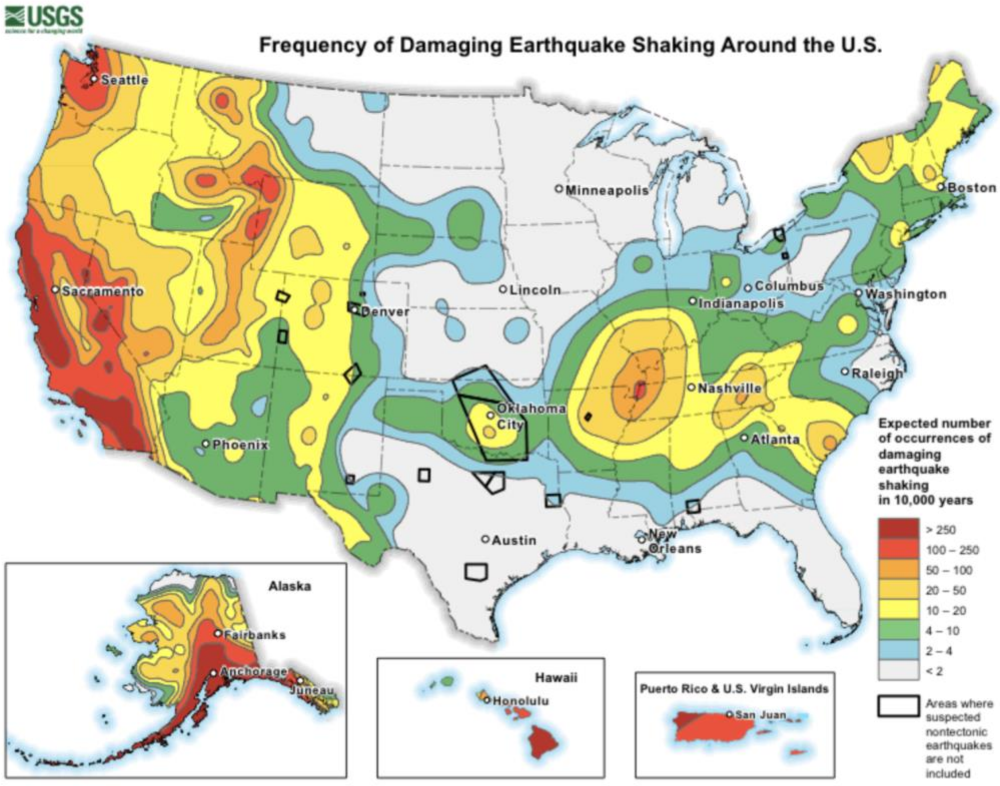


Figure 3: Map showing the expected number of instances of damaging earthquake shaking over a 10,000 year period. https://www.usgs.gov/natural-hazards/earthquake-hazards/science/introduction-national-seismic-hazard-maps?qt-science_center_objects=0#qt-science_center_objects

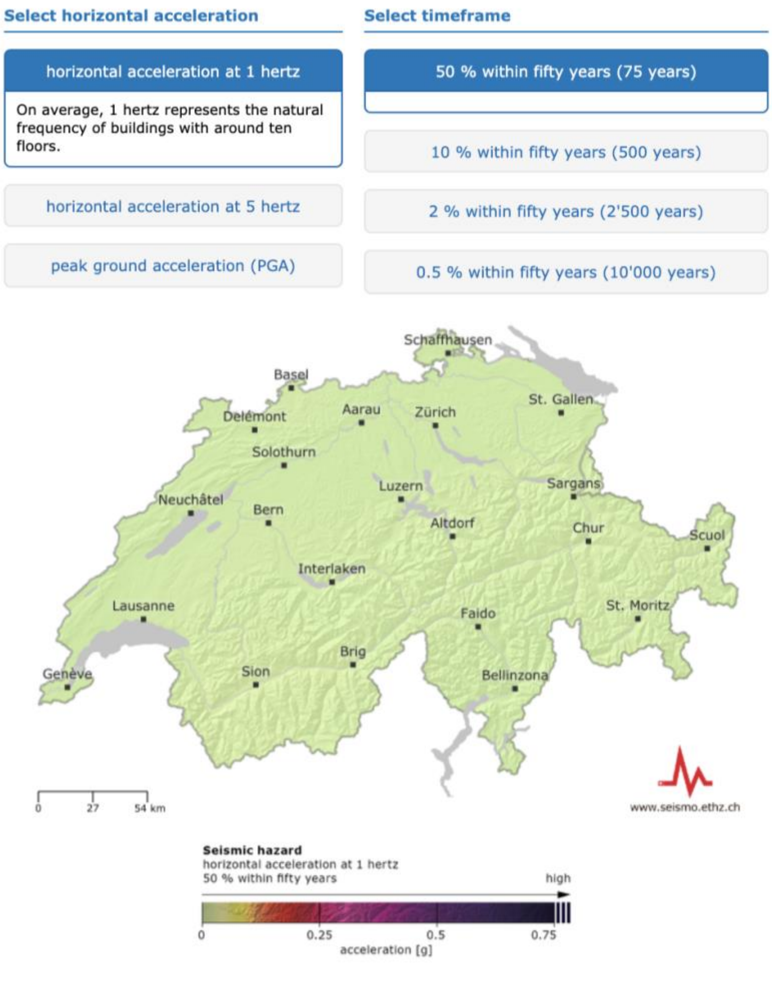
The Swiss Seismological Service (SED), Switzerland

The Swiss Seismological Service (SED) provides three different maps to communicate how likely earthquakes are in Switzerland:

1) Hazard maps: These show how often buildings are affected by particular incidents of horizontal acceleration.
2) Effects maps: These focus on the likely consequences of an earthquake.
3) Magnitude maps: These show how often earthquakes of a particular strength occur.

Hazard maps were designed based on best practices and tested with different target audiences. The detailed results of the testing are reported in (Marti et al., 2019).

The maps can be found here: http://seismo.ethz.ch/en/knowledge/seismic-hazard-switzerland/maps/hazard/ and see Figure 4 for an example.



Seismic hazard map: horizontal acceleration at 1 hertz; the probability of a building constructed on rocky subsoil experiencing this is 50 % within fifty years (75 years).

Figure 4: Example Swiss hazard map

Evaluation of seismic hazard maps in the literature

Expert-based insights into the design of seismic hazard maps can be very useful, however some features common to map design have been maintained simply due to their being the status quo, and may not always be effective tools for communication. An example is the use of proportional circles on a map to represent different magnitudes; this is common practice in cartography but research on human perception has shown that people find it hard to accurately assess volumes and areas (where there are multiple dimensions to be considered) as compared to assessing lines of different lengths (where there is just one dimension) (I.M. Lipkus & Hollands, 1999). This highlights the need for evaluation in the design of seismic hazard maps to ensure they are comprehensible, trustworthy and actionable.

Unfortunately, there appears to be a lack of empirical evaluation of seismic hazard maps in the literature, excepting the work of (Marti et al., 2019)). These authors took a mixed method approach to evaluating comprehension of the three map types (hazard map, magnitude map and effects map - see above) used to communicate seismic information by the Swiss

Seismological Service (SED) in a variety of non-expert users, including the general public and architects and engineers who do not specialise in seismic retrofitting. The SED hazard maps had been designed based on best practice from the hazard visualisation literature wherever possible, including ensuring that legends were visually prominent and contained both qualitative and numeric information, and using darker colours on the map to depict higher hazard areas (Gaspar-Escribano & Iturrioz, 2011; Marti et al., 2019). (Marti et al., 2019) demonstrated that non-expert users could effectively distinguish between hazardous and less hazardous areas when using these maps (although comprehension was significantly and positively related to participant numeracy). Their participants, however, were less able to properly interpret magnitude and effects maps. The authors suggested this may be due to the fact that, unlike the hazard maps, the design of the magnitude and effects maps had not been so closely based on best practice. For example these maps had low contrast ratios, which are thought to reduce comprehension and readability (Hagemeier-Klose & Wagner, 2009; Kunz, Grêt-Regamey, & Hurni, 2011; Marti et al., 2019), concluding that a redesign might be necessary and that including users in this design process could be useful.

(Marti et al., 2019) went on to examine participants' comprehension of statistical information relating to the maps. They found that 73.3% of participants could correctly interpret that a statement that described an earthquake event as occurring "within" a particular period of time meant that the earthquake could occur at any point during that period. Once again comprehension was positively related to numerical ability. They further tested the effect of map interactivity on comprehension in their smaller participant sample of engineers and architects, but concluded there was no effect in this instance.

Operational Earthquake Forecasting

What are Operational Earthquake Forecasts?

Operational Earthquake Forecasts (OEFs) are location specific, time-dependent probabilistic forecasts of the occurrence of earthquakes. As the product of models that combine time-independent modelling based on fault and historical data with time-dependent localised earthquake clustering models (M. C. Gerstenberger, Wiemer, Jones, & Reasenberg, 2005), OEFs provide real-time information on the changes in likelihood of earthquake events across time and space (e.g. (Field et al., 2016; M. Gerstenberger, McVerry, Rhoades, & Stirling, 2014; T. H. Jordan, Marzocchi, Michael, & Gerstenberger, 2014; Thomas H. Jordan et al., 2011; Marzocchi, Taroni, & Falcone, 2017).

The practical usefulness of OEF has been questioned, often based on whether the low probability forecasts and/or probability gains that are typical of OEFs can be of assistance to a risk manager on the ground (Peresan, Kossobokov, & Panza, 2012; Wang & Rogers, 2014). While it is indeed unlikely that an evacuation order could justifiably be given on the changes in probability typical of OEFs under most circumstances, arguably there are many smaller actions that could be taken if the relative risk of an earthquake is raised, even if the absolute risks remain small. Rehearsing emergency drills, ensuring back-up power and water supplies are on hand, and that communication chains are in place are all low-cost activities which could make a big difference to any response to a seismic event. The probability increases observed during seismic sequences in the wake of earthquakes are much larger, and could be used to inform higher-cost decisions such as whether or not it is deemed safe for people to return to buildings. Furthermore, there are other practical uses of OEFs, as well as ethical reasons why any information about increased likelihood of an event should be (carefully) communicated.

In thinking about these uses, it is valuable to clarify that 'everyday' forecast communications, such as OEFs, are very different from emergency warnings; whilst an emergency warning is there to trigger behaviour in acute circumstances, forecasts are there to provide regular information that can inform how a situation might be changing, and upon which low-cost behaviours may be taken based on this information should the individual decision maker see fit. In addition, the regular communication of such information ensures that channels of communication are kept constantly open, and formats of information provision familiar to their audiences. Their success might thus be measured as whether an individual or organisation has higher levels of comprehension about relevant seismic activity, and hence a better basis of information on which to take decisions after receiving OEF information, rather than whether an individual has changed their attitude or undertaken a particular behaviour.

On this basis, OEFs have the potential to be incredibly useful, both operationally and to individual members of the public, and have been evidenced as such during the Canterbury and Cook Strait earthquake sequences in New Zealand for example ((M. Gerstenberger et al., 2014)). The occurrence of prolonged aftershock sequences such as Ridgecrest (California) or the Canterbury Earthquake Sequence (New Zealand) highlights the necessity of communicating this dynamically changing background hazard level to inform risk managers (Becker et al., 2019). Such information can then be used to inform whether or not to rehearse disaster response drills, ready teams of emergency responders, advocate for organisational and household preparedness, time the demolition or repair of buildings during an aftershock sequence, and even inform individual level responses such as checking emergency supply kits or ensuring household items are properly fixed to walls (e.g. (Goltz, 2015; T. H. Jordan et al., 2014; Woo & Marzocchi, 2014); (McBride, Llenos, Page, & Van Der Elst, 2019); (Becker et al., 2020, 2019; Becker, Wein, Potter, Emma Doyle, & Ratliff, 2015)). To effectively inform these decisions however, OEF have to be communicated in a clear, comprehensible and trustworthy way.

One of the few countries that has issued short term earthquake forecast is New Zealand. The National Geological Hazards Monitoring Network or “GeoNet”, releases long-term seismic hazard information in a table accompanied by written information that provides the average number of earthquakes and the probability of one or more earthquakes expected to occur within different time frames. The chosen time-frame depends on the level of seismic activity. During a period of quiescent activity, longer time frames (e.g. a window of 365 days) is used. During a period of elevated activity, shorter time frame (days to months) Operational Earthquake Forecasts are issued. See Figure 5.

	M5.0-5.9			**M6.0-6.9**			**M≥7.0**		
	Average number	**Range ***	**Probability of 1 or more**	**Average number**	**Range ***	**Probability of 1 or more**	**Average number**	**Range ***	**Probability of 1 or more**
within 30 days	1.0-1.4	0-4	60-75%	0.10-0.13	0-1	9-12%	0.01	0-1	1%
within 365 days	6-7	2-13	>99%	0.6-0.7	0-3	40-50%	0.06-0.07	0-1	6-7%

*Calculated for 23 May 2021, 6 pm. This table shows a forecast for future earthquakes in the given time intervals from the time of issuing for the forecast area from 178.50-181.5 degrees longitude and -36.5 to -38.5 degrees latitude. *95% confidence bounds*

For example, this table says that:

- It is about as likely as not (40-50% chance) that there will be one or more M6.0-6.9 earthquakes within the next 365 days.
- It is very unlikely (1% chance) that there will be one or more earthquakes of M7 or above within the next 30 days.
- While the probability of M7 or above is small, the possible impacts of an event such as this could be severe.

Figure 5: New Zealand Geonet Operational Earthquake Forecast showing the range, average and likelihood of earthquakes of magnitude ranging from respectively 5-5.9, 6-6.9 and 7-7.9 over two different time frames (within the next 30 days and within the next 365 days). https://www.geonet.org.nz/earthquake/forecast/

Challenges to communicating OEF

There are several challenges to achieving clear, comprehensible and trustworthy communications of OEF. Firstly, OEFs are, by their nature, probabilistic forecasts and not concrete predictions, which immediately introduces uncertainty into their communication. Since people experience an aversion to ambiguity they often try to avoid uncertainty (C. Camerer & Weber, 1992; Keren & Gerritsen, 1999), which may lead to a rejection of OEF forecasts by the public. It is encouraging to note that the public do have a natural expectation of uncertainty however, at least about future events, and there is evidence that they are willing to accept probabilistic forecasts of the weather, for example (Joslyn & Savelli, 2010; Morss et al., 2008). However, whether they are tolerant of the much higher degree of uncertainty around OEFs and the lack of public familiarity with these types of earthquake forecasts still needs to be evaluated.

Secondly, these probabilities can vary rapidly over several orders of magnitude in time and space. Visualisation can be a very useful way of communicating numbers such as those from OEFs, however designing a visualisation that can display the huge variation in possible forecasts is challenging. Finally, most of the time the probabilities outputted from OEF forecasts are very small, and people struggle to understand these small numbers ((C. F. Camerer & Kunreuther, 1989; Halpern, Blackman, & Salzman, 1989; Isaac M. Lipkus, 2007)) and compare between them (Cohen, Ferrell, & Johnson, 2002; Kaplan, Hammel, & Schimmel, 1985).

These challenges combine with those from the psychology of human numeric perception - different presentations of the same number can make it feel very different to an audience, changing their perception of the risk (e.g. (Denes-Raj, Epstein, & Cole, 1995; Freeman et al., 2021; Pighin et al., 2011; Yamagishi, 1997). Choosing to display the number of earthquakes in a 100 year period vs the percentage chance of an earthquake in this same time period, or communicating an absolute vs a relative probability for example, can have real perceptual and behavioural consequences. In turn, numbers need context to give them meaning to people, particularly small numbers (Spiegelhalter, 2017); is a 1% chance of a magnitude 5 earthquake within the next 7 days high or low? Compared to what? And that choice of context can fundamentally alter people's perception of the risk (Freeman et al., 2021; Keller, 2011; Sandman, Weinstein, & Miller, 1994b). Communicators then, have quite a lot of power to change people's perceptions of a number simply by changing the way that number is presented, thus this power must be used judiciously if they and their associated communications are to be trustworthy.

In an attempt to address these challenges in our design of OEF communications, we used insights from the literature and inspiration from other fields to produce a baseline design, from which we have iterated in a co-design process with target audiences in Italy, Switzerland and Iceland, taking into account their needs and integrating their feedback at each stage in the development process. Here, we present the results of this process, culminating in the most recent design so far and the steps we have planned to continue to take the process of design and evaluation forward.

Dashboard

After initial qualitative interviews with participants across Italy, Switzerland and Iceland discussing some early ideas for ways to represent earthquake forecasts, our first full design took the form of a dashboard, which provided a graphical way of bringing different formats of dynamically changing data together visually, on one display, thus allowing patterns and anomalies to be quickly identified (Brath & Peters, 2004; Few, 2006). Using a dashboard allowed us to reduce the influence of perceptual biases created by any one particular format (for example, relative risks making small absolute differences look large). The dashboard also allowed us space to provide various types of contextual information to help users make meaning of the central OEF statistics. We produced a version typical of a forecast for a single location that might be seen during a period of quiescence, where there is little earthquake activity (Figure 6) and one typical of what might be seen during periods of elevated seismic activity in the wake of a damaging earthquake (

Figure 7). Each component of these dashboards was designed to be able to be taken out of context of the dashboard without losing important information.

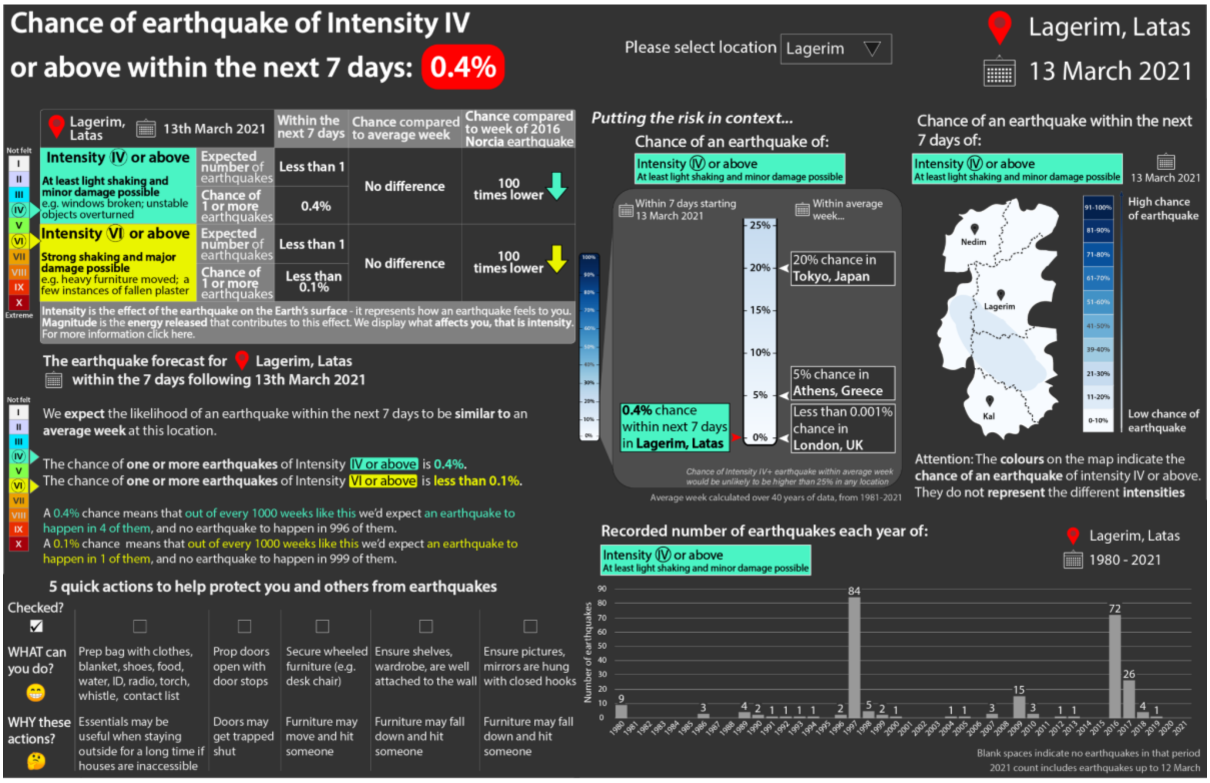


Figure 6: First iteration of a potential OEF dashboard design showing a hypothetical location during a quiescent period

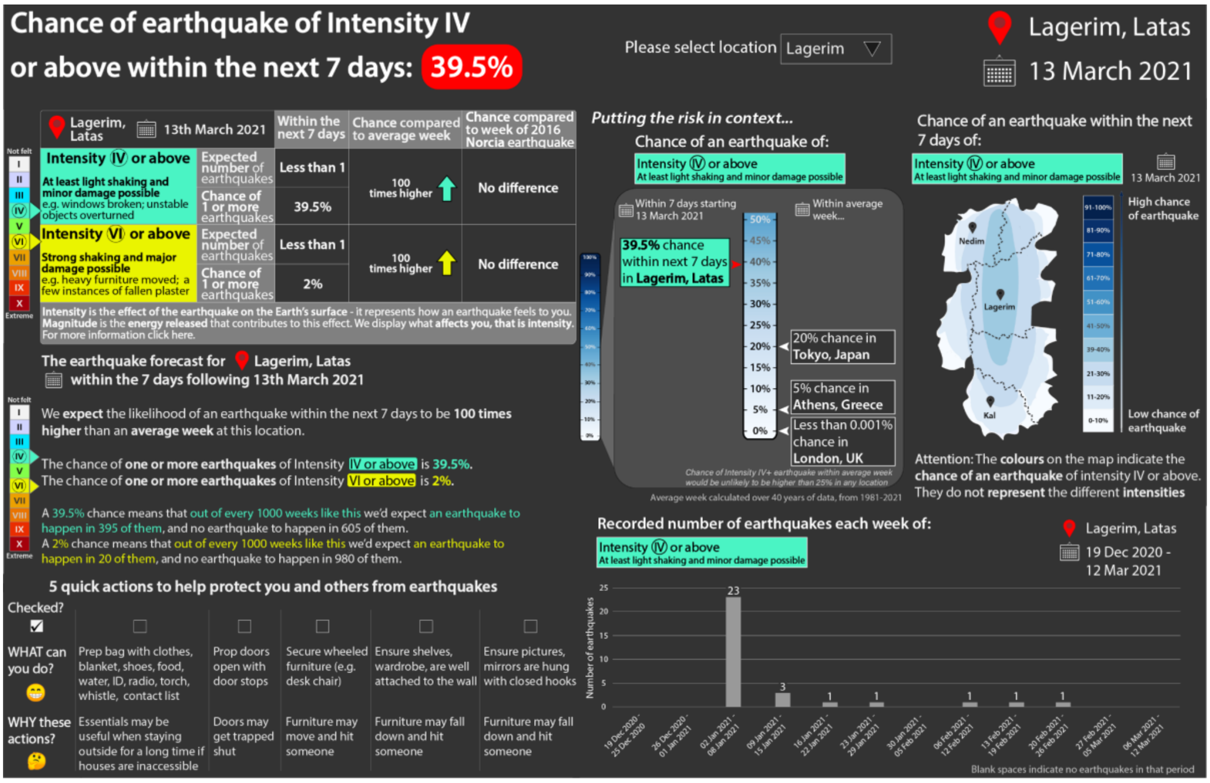


Figure 7: First iteration of a potential OEF dashboard design showing a hypothetical location during an active earthquake sequence

In designing this dashboard, key decisions had to be made regarding its overall characteristics before any of its individual components could be designed. The most far-reaching of these was whether to communicate probabilities relating to the magnitude of an earthquake (one of a variety of measurements of the size, relating to either the length, width and slip involved, or the energy released at the earthquake's source) which gives a single figure for each seismic event, or relating to the intensity of the earthquake (the strength of shaking as a result of the earthquake) which will vary depending on the location of the observer compared with the earthquake (including the depth of the earthquake and the type of geology).

Proponents of communicating one of the measures of magnitude suggest that it is more familiar to the general public (and indeed most earthquake communications outside of Japan relate to magnitude rather than intensity). As such it may be less likely to be both misunderstood and/or rejected due to novelty. However, there are several downsides of basing forecast probability communications on magnitude too. Firstly, because of the very large range in magnitudes of different earthquakes, a non-linear scale is used, allowing this variation to be collapsed into a visually manageable scale. However there is evidence that people do not comprehend non-linear scales very well and thus such scales may modify individuals' risk perceptions in disproportionate ways (e.g. (Heckler, Mikula, & Rosenblatt, 2013; Menge et al., 2018; Romano, Sotis, Dominioni, & Guidi, 2020) An increase of one point on the Me (energy magnitude) scale is equivalent to around a thirty fold increase in energy release, whilst an increase of two points is a 1000-fold increase in energy (Hayes, n.d.). As such, although a magnitude 7 earthquake is "only" 5 points higher on the scale than a magnitude 3 earthquake, it releases a million times the amount of energy. The other concern with communicating magnitude is that it is a measure of the size of the earthquake itself, but not its effects. How a seismic event affects a nearby city is contingent upon the depth of the earthquake and its distance from the city, as well as characteristics such as the local geology and the seismic resistance (vulnerability) of the building stock. Thus, a magnitude 5 earthquake may cause limited damage in Iceland where buildings are broadly seismically resistant, but have quite catastrophic effects in Italy.

Proponents of communicating intensity suggest that it is more representative of what an individual will experience at their specific location, and the damage that might be done. There are also downsides to communicating intensity, though. People tend to be less familiar with intensity as a measure than they are with magnitude, although this could reduce over time if intensity were more frequently communicated. However, because of its location-specificity, an earthquake does not have 'an intensity'. The range of possible intensities that can be experienced for any one particular magnitude earthquake can vary substantially over even short distances such as between adjacent buildings. Consequently, unless the resolution of an intensity forecast is very high, the relevance of that forecast to an individual on the ground may be limited to some extent. It also means that intensity is never likely to be used in the broadcast media when describing an earthquake, meaning that the term is unlikely to become familiar in public discourse.

For our initial dashboard design, we decided to communicate probabilities relating to earthquake intensity rather than magnitude, based on it being more representative of people's experience of an earthquake at a particular location. Even after this decision had been made however, we needed to decide what threshold(s) to communicate these forecasts over. Since

Intensity IV is that from which earthquakes are typically felt, we settled on a threshold of Intensity IV and above, on the basis that audiences are in general more tolerant of ‘false alerts’ than apparently ‘missed alerts’ (discussed in Brooks et al., 2021)).

The dashboard components

The dashboard was initially designed to be shown to people as a non-interactive page, and comprised six different components:

1. Table - a familiar presentation format, decreasing the chance of rejection of the dashboard due to novelty. It provides detailed forecast information for two levels of earthquake: Intensity IV+ (noticeable) and Intensity VI+ (damaging) earthquakes within the time frame of the next week. More specifically, the table provides information on the expected number of earthquakes of each intensity within the next week and the absolute probability of one or more earthquakes. It also provides two ‘relative risks’, designed to give context for the central forecast statistics. One was the probability gain compared to the location’s “average week”, allowing users to see if the forecast in question was “high” or “low” compared to what is the “norm” for the location. The other was the probability gain compared to a (high likelihood) forecast immediately after a memorable event (the 2016 Norcia earthquake) to again provide users with a sense of whether the forecast in question was high or low.

2. Text - an explanation of the major numeric details from the table as an alternative format, reinforcing and reframing the information in the table, and catering to the needs of a broader variety of users across varying numerical abilities. For example: “A 0.4% chance means that out of every 1000 weeks like this we’d expect an earthquake to happen in four of them, and no earthquake to occur in 996 of them.”

3. Risk ladder (or ‘thermometer’) - designed to provide context to the forecast numbers by showing where that numerical likelihood lies on a scale compared with other risks which might be familiar and comparable. On one side of the scale is shown the forecasted probability of an Intensity IV+ earthquake within the next 7 days in one specific location, and on the other side, the probability of earthquakes at or above that same intensity threshold in an average week for different familiar cities. The idea is that people have intuitive mental models for what is high and low risk for a particular event (Granger Morgan, Fischoff, Bostrom, & Atman, 2002). For example Tokyo might be considered high risk for some people when they think of earthquakes, whilst London might be considered low risk. Knowing where one’s own city lies between the two might help people attach a ‘feeling’ to the likelihood. Users do not need to know the specific probabilities associated with the comparator cities, but they need to be chosen such that they are familiar enough that the audience have an intuitive sense of whether they are high or low hazard areas. The choice of comparator can have a marked effect on users’ perceptions of a risk and thus it is a challenge to choose comparators that are informative and not persuasive (Freeman et al., 2021; Roth, Morgan, Fischhoff, Lave, & Bostrom, 1990). Working with the audience to refine the choice of appropriate comparators is thus essential (Freeman et al., 2021).

4. Map - designed to provide geographic context for the location's forecast by showing how the forecast varies from location to location within the focus country. Similar to the risk ladder, this serves as a form of contextualisation for the forecast, but at a more local geographic scale, and one that is visualised spatially.

5. Bar chart – Whereas the map aimed to provide geographic context for the forecast, the probability of an earthquake of course varies in time as well as in space. The bar chart was designed to provide this temporal context by showing how the present forecast compares to previous recorded earthquake activity in the same location. In our hypothetical example from quiescent times, the chart showed the recorded number of earthquakes in the location each preceding year. In our hypothetical active period, we used it to show the number per week for the preceding 12 weeks. Just as with the choice of comparators in other components of the dashboard, the choice of size of area considered and time frame (both the length of time visualised and the categories into which the timeframe is "binned") can affect people's perceptions of the information.

6. Actions people can take - a question audiences often ask of risk communications is "What should I do in response to this?". This list of low-cost, easy actions was designed to answer that question. Whether it belongs with the forecast depends on the legal and ethical framework within which the OEF is being produced.

User centred design - focus groups

Eleven members of the Italian public were invited to participate in one of four virtual focus groups online, designed to gain insights into their understanding and emotions relating to the dashboard and each of its individual components. During these interviews, participants were shown the overall dashboard (translated into Italian) and then each individual component and asked a series of questions probing their emotions, impressions and understanding of the information, including what they find useful, trustable, actionable and worth receiving. Additionally, six seismologists from various institutions were invited to participate in a further two virtual focus groups. They were asked similar questions to the public participants, although some more technical aspects of the data and dashboard were also discussed. Insights from all these focus groups regarding the overall dashboard design and each individual component are summarised below.

Overall dashboard/General themes

- **Information overload**

The general consensus across both public and seismologist participants was that the dashboard was too visually cluttered, making it difficult to extract the key take-home messages and navigate to the most important components. Indeed, different participants noticed different things on first glance, further emphasising the lack of visual focus in the design.

Participant in public focus group 1: "I start to ask myself, but where am I, what is the forecast?"

It was suggested by both public and seismologist participants that the interface could be simplified by making it interactive, keeping only the key take-home numbers and information on the front page, and using clickable layers to provide other components that are perhaps more for particularly interested users. Public participants also suggested removing certain components they deemed as irrelevant.

Participant in public focus group 1: "I would try to distinguish the topics according to their importance... the order is consequential, so once I was informed about the forecast, then later I looked for information about the history of the earthquakes."

It was clear from all the feedback that the interface of the communication needed to be greatly simplified, moving away from the concept of a dashboard (which purposely combines multiple visualisations on the one page), and instead making use of layout, colour, font size and layering to assist users in quickly identifying the simple take-home messages (getting the "gist"). Nevertheless, there was some indication that the requests for the layering or removal of some dashboard components may have in part been due to misunderstanding of the broader purpose of the dashboard and of these individual components, and better labelling of components could make clear what perspective they were designed to give.

- **Purpose of the dashboard and its individual components**

There were several parts of the discourse in both public and seismologist focus groups that highlighted misunderstandings regarding the purpose of the dashboard (and of OEF more generally). In the focus groups, participants were specifically not told the aim or purpose of the dashboard in order to simulate them coming across it online.

As discussed earlier, the aim of OEF is to provide regular information that can inform a range of potential audiences how a local seismic situation might be changing. Decision makers (such as individual members of the public or risk managers) may use such information to inform (typically low cost) actions, however some participants appeared to view the dashboard as a tool to direct behaviour in acute situations during or immediately after an earthquake (the only seismic information they are currently familiar with), and thus questioned the value of some of the more contextually motivated components of the dashboard, and why it was not telling them precisely what to do in an emergency situation.

Participant in public focus group 1: "I thought [the dashboard] was made to alert people. Curiosity is ok but I am not sure why [the bar chart] is inserted into the project you are working on."

Another misunderstanding was about the capabilities of OEF. There was an indication from some of the public participants that they expected OEF to provide precise predictions (or at least high probability forecasts) of when and where an earthquake would occur. For example, some discussed the information in the dashboard allowing them to take certain actions such as sleeping in the car or moving to a different region during time periods when an earthquake is forecasted.

Participant in public focus group 2: "It would be nice to have a monthly calendar because if you tell me that within the next week you expect less than one earthquake, it means that you do expect an earthquake but maybe not within the next week, so if you give me the month, I don't know, a widget of the month with a red week, a yellow week, a green week, if you give me an opinion that is updated in time, I know which week I should be more careful, so if I need to choose whether or not to sleep in the car, I choose the initial week instead of the last one."

Interestingly, this misperception also led to doubts about the reliability of the dashboard in some participants who were aware that earthquakes could not be predicted and thus who doubted the perceived "predictions" the dashboard was making.

In turn, the purpose we intended for each individual component, and how we'd hoped it might aid understanding/be useful, was misunderstood by some participants in both the public and seismologist groups.

Participant in seismologist focus group 2: "With this [bar chart] I do not know what information I need to understand. There are 23 earthquakes in January, but I don't know what I can do with this information."

Participant from public focus group 2: "Knowing the past [earthquakes], I will not know the future ones. It is simply data that is interesting but it does not tell you anything about the future."

These component-specific misunderstandings are discussed in more detail in their respective sections below.

Clearly, communication of such an unfamiliar topic as an earthquake forecast needs to include more explanatory information to make the purpose, uses and limitations of a dashboard and its individual components much clearer to users - particularly that it does not provide an alert nor a precise prediction of the time and place of an earthquake. Options might be to have more introductory text, which itself could clutter a page even more, or optional information such as clickable explanations for each component, a scrolling walkthrough, or a video guide - with the caveat that such optional features are often ignored by most users.

- **The probabilistic nature of forecasts**

Combined with misapprehensions about the purpose of the dashboard, or a forecast more generally, was public participants' discomfort and unfamiliarity with the concept of probability or likelihood compared to being told categorically what was going to happen and what actions they should take (i.e. a prediction, accompanied by persuasive/alert rather than informative communication). This was despite us mentioning weather forecasts in our introduction as a way to find the closest possible analogy - especially as weather forecasts are increasingly expressed in probabilistic terms.

It became clear from some comments that many people translate - or want help translating - probabilities given in weather forecasts into a categorical 'yes' or 'no'.

Participant in public focus group 1: "It is implied in the concept of forecasting the idea that they can be false, and if the likelihood is 80% you think that tomorrow it will surely rain."

And for many the concept of probability is so alien that they essentially ignore it:

Participant in public focus group 1: "What I am saying is that I don't care about the likelihood. This morning a likelihood was given and it is not raining. I want to know the intensity [of the rain forecast] as I want to know the millimetre of water. Because the percentage then can say anything or nothing. It is of little use"

For some members of the public, this difficulty in knowing how to interpret and use forecast information made them even doubt its usefulness at all:

Participant in public focus group 1: "According to me, playing with likelihoods that are not definitive, it is better to have other information, not how likely it is to happen, but if it does happen, what can happen?"

And some had felt their trust in probabilistic information had been undermined by the miscommunications around the L'Aquila earthquake:

Participant in public focus group 1: "On this matter of the probability, it is similar to what happened in L'Aquila... in L'Aquila there were so many earthquakes, the final weeks before 6th April were weeks of nightmares, and we were continuously told to stay calm... we have *been reassured on the basis of a statistical forecast, completely wrongly - and we trusted it, and so what evaluation should I make of this graph?"*

All these difficulties highlight the challenge that operational earthquake forecasts present to communicators. The instinct for every human is to want certainty, but that is an instinct that cannot be sated without risk. Converting a chance to a certainty ('it is going to happen' or 'it is not going to happen') creates a risk of being wrong - and, as shown by the events of L'Aquila, that carries heavy consequences. Instead, forecasters can try to communicate the concept of 'this might happen' along with the appropriate sense of 'how likely' it is.

However, to do this successfully requires priming of the audience's expectations, as discussed in the section above, and which we had (unsuccessfully) attempted to do by using the familiar word and concept of 'chance' instead of 'probability' or 'likelihood' throughout. It's clear from the audience's comments that they also want as much assistance as possible to help interpret the unfamiliar probability information. We cannot give them the certainty they crave, but we can help them get a 'feel' for the risks.

- **Giving context to the probabilities in the forecast**

The table, text and risk ladder formats try to lend context to the absolute probability of an event calculated in the forecast. Giving context is a well-known necessity in risk communication, as a probability on its own is very difficult for people to interpret: without existing knowledge of the range of probabilities it is often difficult to assess how high or low one feels a probability to be.

Using expected frequencies (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007) is one way to give context - turning a probability into the number of times an event would be expected to happen in 100 (or 1000) identical situations. This, however, is only really applicable in a scenario where the identical situations are easily imagined (e.g. 'out of 100 patients like you, we'd expect...'). For very low probabilities and natural hazards it is far less easy to visualise the situation as it moves into the truly hypothetical (e.g. 'out of 1 million possible future versions of next week, we'd expect...'). We explored this possibility in the text format of the dashboard, and the reactions to it are discussed in the 'text' section below.

An alternative method of giving context is to help position the risk relative to other risks. This may be literally in the form of a numerical relative risk (e.g. '100 times higher than average'), or by giving comparison risks - often graphically marked along a number line, called a risk ladder.

Past research has typically used comparator risks from different domains, for example comparing the risk in question to the chance of being in a car accident, or getting struck by lightning, or dying from cancer. In the media it is very common for communicators to look for other risks which have a similar likelihood of occurring, although they may differ dramatically in impact as well as other factors that affect the emotional component of the risk (Slovic, Fischoff, & Lichtenstein, 1981).

There is a paucity of work evaluating the usefulness and impact of these kinds of comparators, given their popularity in risk communication. Recent work we undertook ourselves, interviewing people about useful context for communications of an individual's personalised risk from COVID-19 (Freeman et al., 2021) found that comparisons with other types of people's risk from COVID - a 'within-the-risk' comparison (so a 'risk persona' such as the "risk for a 90 year old with diabetes, kidney disease and Parkinson's"), were found to be considered more useful than cross-domain, between-risk comparators (such as the risk of dying from cancer). There is also an unfortunate history of between-risk comparisons being used to try to persuade people that one risk should be 'acceptable' to them if it has a similar likelihood (or even consequence) as another risk - which ignores the fact that risks have many different emotional dimensions (Slovic et al., 1981), although the topic of which comparators to use has been under debate for decades (Slovic, Kraus, & Covello, 1990).

Bearing our own previous work in mind, we trialled similar, within-the-risk comparators in the dashboard, providing context to the earthquake forecast by comparing it to other earthquake likelihoods in the table and risk ladder formats of the dashboard. Feedback was sought on the perceived usefulness of these within-the-risk earthquake comparators from focus group participants.

Discussions amongst public participants were largely related to the perceived value of the specific comparator earthquakes (Norcia) and cities (Tokyo, London, Athens) - results are reported in the table and risk ladder sections respectively. Whilst this content was also discussed amongst seismologist participants, there was also discussion about the idea of using risks other than earthquakes to provide context to the forecast, with some preferring the idea of using cross-domain risks:

Participant in seismologist focus group 2: "The comparison could be with something different from an earthquake that can still happen within the next 7 days, something that can be related to something that everyone has had experience with."

Another suggestion was to create categories of likelihood (e.g. 0-10%, 10-20% etc) and add to each a (cross-domain) comparator risk, ensuring that these risks were emotionally matched to that of an earthquake happening (i.e. a negative emotion, and presumably also a risk with similar positions on the dread dimensions (Slovic et al., 1981))

In this round of interviews, questions weren't asked specifically of the audience members whether they found within- or between-domain risk comparators more helpful. Further research will help clarify this.

- **Communicating magnitude vs communicating intensity**

Among seismologist participants, magnitude was considered to be more familiar to the public, and simpler in the sense of there being only one possible magnitude per earthquake, unlike intensity, for which each earthquake is a continuum from a maximum near its epicentre down to zero for those far away from it. Nevertheless, it was noted that intensity is more representative of the individual's experience and thus could be more useful to them, provided they don't dismiss it on grounds of unfamiliarity. There was discussion about possibly communicating both magnitude and intensity, or allowing users to choose between the two, although concerns were expressed regarding public users mixing them up. It is also possible that due to their familiarity with magnitude, wherever intensity is communicated public users will mistakenly assume it is magnitude information instead. Ultimately, there was no clear consensus either way as to whether to communicate magnitude or intensity.

Participant in seismologist focus group 2: "When you read the newspaper, people are always saying "there was a magnitude such and such..." so I think the general public would be more impressed by reading something about the probability for magnitude, but on the other hand I agree that this information is not the most useful. I think intensity information is more useful but if they do not read the information because they don't care about intensity because they do not know what it is, then of course it is more useful to communicate magnitude, this means that people would read the information in that case."

Among the public participants, again there was no clear consensus between which of the two to communicate, and again the idea of communicating both was discussed. It is worth noting however, that communication of both together would add further detail to an already complicated display, and if layering were used as an approach to combat this, one returns to having to make the decision about which is the primary measure that would be presented in the top layer. In turn, if you allow people to choose, they may choose magnitude over intensity purely on the basis of familiarity. Related to this, one participant noted that given people's familiarity with magnitude, intensity-based forecasts that include magnitude as well may allow people to better calibrate their expectations of what is to come based on their past experience with different earthquakes of different magnitude.

Participant in public focus group 2: "I would communicate both intensity and magnitude so that if you are a person with a prior experience who thinks, with a magnitude 5 I went into the middle of the street, so with a 4 I won't remain at home, even if you tell me to remain at home."

It is worth noting though, that such calibration may be difficult when based on magnitude, due to the large variation in how an earthquake can be experienced in terms of ground shaking. One consequence of this is that a "high risk" forecast might be ignored on the basis of an individual expecting to have the same (minimally disruptive) experience as with a previous earthquake of the same magnitude that they experienced as low intensity, or were in a robust built environment for.

Similar to the seismologist participants, some public participants expressed concerns about intensity and magnitude being mixed up if they were communicated together, and it was suggested that each would need to be carefully defined to try to avoid such confusion, perhaps by providing clickable definitions in a prominent place such that it is clear to the user which of the two they are looking at and how they differ.

Interestingly, some preference for magnitude was based on perceptions of intensity being more subjective.

Participant in public focus group 1: "How I feel the earthquake is subjective data... I would stick to objective data, that is that related to the magnitude."

Whilst it is correct that intensity can be much more variable than magnitude, since it depends on the individual's particular location, it isn't a subjective measure in the scientific sense of the word - it is based on objectively measured parameters such as distance of the location from the earthquake hypocentre and geology of the local subsurface. It was not clear whether participants were using subjective in a colloquial sense, meaning it depends on the individual's location, or whether they meant that intensity is a less scientific measure (or both). Regardless, even though there is no clear consensus from the focus group data regarding whether magnitude or intensity (or both) would be more helpful, a key take home is that each must be clearly defined. In particular, the definition of intensity we used ("Intensity is the effects of the earthquake on the earth's surface - it represents how an earthquake feels to you") could be adapted to avoid terminology regarding feelings and experience, and perhaps simplified to something like "Intensity is the effects of the earthquake on the earth's surface at a particular location". A participant in one of the seismologist focus groups specifically cautioned against the use of experiential terms for reasons of misperception:

Participant in seismologist focus group 1: "I do not like the term "experience" because then people think that intensity is subjective, when it is [actually] an objective measure that is reflected in your experience."

It should also be noted that the official Mercalli Intensity scale descriptions are imprecise (e.g. "Damage negligible in buildings of good design and construction" - what does 'good design' mean?), and it is worth considering the trade-off between their comprehensibility by the lay public versus their possible conveying of intensity as being vague and unscientific. Indeed, some participants in the public focus groups did comment about the vagueness of these descriptions. For example, when discussing the description of an intensity IV earthquake that we provided on most dashboard components, one public participant noted that "minor damage" was vague, and in fact could be hugely variable depending on the location in which the earthquake occurs.

Participant in public focus group 2: "This is very general, there should be something about how the city or the state is prepared… because minor damage possible can mean anything and nothing, if a shed collapses it is different from if a skyscraper collapses. For example 'Houses over 30 years old are at risk'. This is more detailed data that gives you more information than 'minor damages possible'."

The Mercalli scale descriptions were also thought, by some participants, to become more precise as higher intensities were defined, which may affect how people perceived the precision of the forecasts.

Participant from Focus group 1: "The higher it is the more felt it is and the description becomes more precise."

In fact, in many cases, public participants commented on how vulnerability was a key component of risk, and that local vulnerabilities could make a big difference to the impact of an earthquake, and were unconvinced that this was appropriately taken account of by a measure of intensity.

Participant in public focus group 2: "It would be useful to see, since this is personalised for each area, since we are in Lagerim, how much the area is prepared because we know that an intensity in Tokyo and an intensity in Naples are two completely different things. So what should I expect from [Lagerim]?"

As might be expected, they also wanted personalised (and certain) information:

Participant in public focus group 1: "The average man wants to know, yes, but what is happening to me? Will I die? What likelihood have I of surviving? Because the ordinary man, not educated, is aware of living in a seismic area. At my house what is going to happen?"

Such specificity, just like such certainty, can never be given. As for the issues around probabilities and likelihoods, the best response to these desires is to give the best estimates available, and then help people interpret and personalise them for themselves. For example, helping people assess the vulnerability of their surrounding built environment and hence be able to instinctively assess the potential impact of an earthquake striking at that particular time, as well as how to react if it should. As another participant said:

Participant in public focus group 2: "My question is, as a citizen, have I to worry or not? If you are in the historical centre, be worried and do this, while if you live in a residential area, or if your building is new and it does respect the antiseismic norms, you can calm down."

We had anticipated that participants might be more familiar with magnitude than intensity, and perhaps might even mistake any intensity labels with magnitude labels (and there were some hints that some might have done this). To try to forestall this, underneath the table we included definitions of both intensity and of magnitude to highlight that they were two different measures, and made it clear that we were showing intensity and not magnitude information within the table. This, however, was not well received by several of the public and seismologist participants, who wondered why magnitude was being defined within the table when no magnitude information was shown anywhere on the dashboard. Some of the public participants wondered whether they had actually missed this information somewhere. So an attempt at clarity may have created more confusion for our test users. The lesson we took from this was not to include explanations that were not strictly relevant to the information displayed.

One more technical issue to resolve when it comes to choosing magnitude over intensity is exactly what probability is being communicated. For calculations of likelihood based on magnitude, is it the probability of the epicentre of an earthquake of a certain magnitude being located in that defined area, or the probability of an earthquake of that magnitude being felt at all in that defined area? For calculations of likelihood based on intensity is it the probability of a seismic event of at least that intensity being felt anywhere in the defined area? These possibilities need to be explored in greater depth in future research, and then – whatever choice is made – communicated clearly.

- **Choice of threshold values to communicate**

Some participants in the seismologist focus groups questioned the choice of intensity IV as a threshold, suggesting that people aren't guaranteed to feel an intensity IV earthquake, and thus there may be a loss of trust when earthquakes are forecast but not felt. Participants in the first seismologist focus group suggested that intensity VII would be a good alternative threshold, as at least in Italy this is when you typically start to see damage in buildings that are not anti-seismic, although they also noted that different thresholds would be useful for different users. This of course would have to be traded off against the fact that the forecast likelihood would be smaller if the intensity threshold is set at a higher level, possibly leading to discounting of the likelihoods (treating them as zero) on this basis.

Participant in seismologist focus group 1: "VII is when earthquakes start having their effects... [but] the choice depends on what the person would use it for. I, as a citizen, would like to know Intensity VII as it is when damage starts, but VI is useful for the civil protection because intensity VI will be perceived."

In the second seismologist focus group it was suggested that different thresholds might be needed depending on the country in question, perhaps using a higher threshold for Iceland compared to Italy, and Italy compared to Switzerland.

In a similar vein, a public participant who had experienced an earthquake felt that intensity IV was too low, and may cause people to dismiss the information on the grounds that they have nothing to worry about (although some others disagreed with this position).

Participant in public focus group 3: "Noticing that it refers to a IV earthquake, I straight away thought that it is light, I am not worried, I read a little, or I read the minimum necessary... If I had read VII, I would have read every single piece of information that was provided to me."

This hints at a broader misunderstanding about the nature of this threshold information - that the forecast is for earthquakes of intensity IV *or above*, i.e. that it includes earthquakes up to the maximum possible intensity, not just the threshold level itself. This misunderstanding was apparent in other sections of the public discussions too. For example, on most of the individual components of the dashboard, we placed a green box (colour coded according to the Mercalli Intensity Scale colours) that described the threshold of the information as "Intensity IV or above. At least light shaking and minor damage possible." When reading this information, some public participants noted that the descriptions "light shaking" and "minor damage" made them feel reassured. It is possible that these participants were anchoring to the words light and minor in the description, and were again discounting the "at least" at the beginning of the description.

Participant in public focus group 2: "Minor damage possible is reassuring... I would add another box or banner with 'at least minor damage possible' [in addition to the 'at least' before the light shaking]."

It is again clear then, that care must be taken with the terminology used in communications. In this case, more care is needed to make clear the fact that the forecast includes earthquakes from intensity IV up to the maximum possible intensity, and thus damage across this range too.

In the Table component of the dashboard, a second threshold of Intensity VI or above was shown, in addition to the main threshold of Intensity IV. This was intended to be a threshold from which earthquakes could be considered damaging. Some public participants however, expressed that VI may be too similar to Intensity IV to be a useful second threshold to reference, and may in turn be too low a threshold to be considered a "damaging" earthquake. Further discussions regarding these two thresholds in some of the public focus groups highlighted again that the concept of the forecast being set at intensity IV *or above* might not be salient to people. Specifically, participants in most of the public focus groups experienced some misunderstanding about what intensities each 'category' (IV or above and VI or above) included, with many thinking that they were non-overlapping. One participant expected the two intensity thresholds to represent the forecast of earthquakes below intensity V and above intensity VI respectively (i.e. two discrete and non-overlapping categories of intensity).

Participant in public focus group 2: "It was strange to me to see intensity IV or above and intensity VI or above, namely I expected to see the entire scale, I don't know, V and below and VI or above. So that a part of the table took the first 5 levels and the other part the other remaining levels"

Several others thought the thresholds were that IV or above contained only intensity IV and V, whilst VI and above contained all the others.

Participant in public focus group 3: "I considered the IV or above only counting the IV and V and the other being VI or above."

As discussed, these misunderstandings may in part be due to poor explanation of the specifics of the intensity thresholds, and precisely what the statement "or above" means in the dashboard design. The colour coding of the threshold information in each component of the dashboard may have contributed further to these misperceptions since we had chosen to match the colour of each box referring to the intensity threshold with the associated Mercalli scale colour for the value at the *lower* end of the threshold. This may have reinforced a focus on this single value. A combination of an improved explanation of "or above", different colour coding or graphical indications of what was included in the forecast, and, for the table, perhaps simplifying it to show only one intensity threshold as a default may solve some of these issues with comprehension.

- **Interactivity**

Some participants in the public and seismologist focus groups wished for the dashboard intensity threshold to be customisable such that they could select a particular intensity and see how the dashboard components and forecast numbers changed, and get the Mercalli definition for each intensity selected. When looking at the table, some public participants mentioned that they would like to be able to find out the equivalent forecasts in "units" of magnitude, to compare with the intensity forecasts (perhaps in part prompted by the definitions of magnitude and intensity beneath the table). There were various other requests for other types of interactivity from several public and seismologist participants, including the aforementioned request for layers of information. As discussed, there is value in the user being able to interact with the dashboard to gain more detailed information, allowing the main interface to be simplified and the take-home messages to be more easily absorbed. However, most participants use the default settings of any website. In this specific case as well, interactivity could be a problem: changing the parameters of the forecast (such as the intensity threshold or area over which the forecast is calculated), as is possible in some interfaces already (e.g. https://www.richterx.com/?go=forecast), provides a myriad possible forecast likelihoods that may overwhelm user attention. It may not be immediately obvious to users why probabilities increase as areas of interest increase, or time periods increase, causing confusion. As such we decided in future iterations to hold as many of the forecast parameters as possible constant, and actually to reduce options, focussing users' attention on a single likelihood and freeing up their attention for other components of the dashboard that help lend context to that forecast.

Feedback on individual components of the dashboard

1) Table

The table (see Figure 8) was designed to show the forecasted likelihood of one or more earthquakes within the next 7 days at two different thresholds (IV+ and VI+), alongside the associated rate of earthquakes forecasted at each threshold. The phrase 'within the next…' was specifically chosen based on research by Doyle et al., 2020, which suggested that this made it clearer that the forecast event could happen at any time during the specified time period, overcoming a natural bias for assuming it was more likely to happen near the end of it. Contextual information was provided by comparing these forecasts to the "average week" for the location in question, and to the forecast probabilities in the week following the M6.2 Norcia earthquake that occurred in Central Italy in 2016.

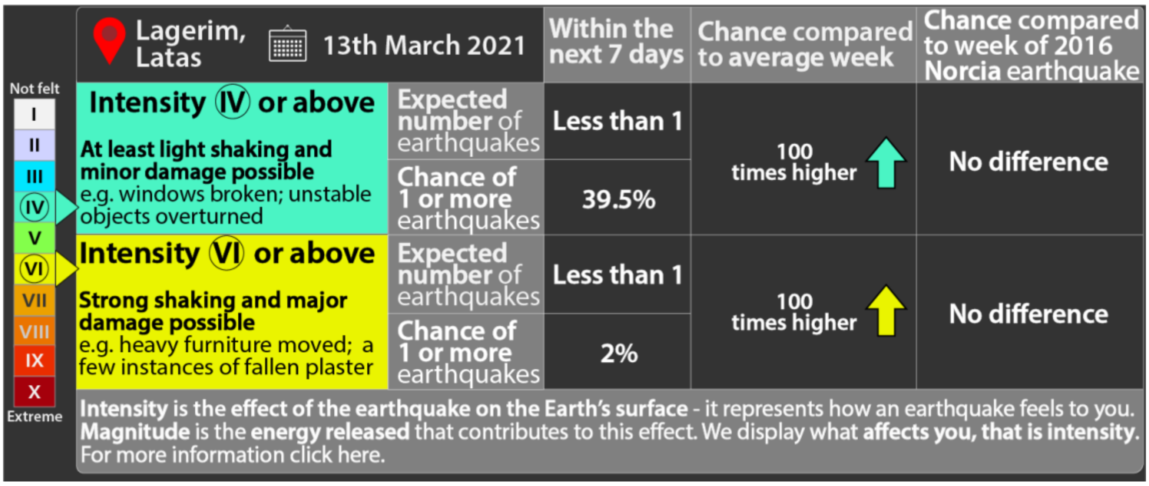


Figure 8: The Table format within the dashboard. Participants were shown either this version, or one in which the columns and rows were reverse to form a 'vertical' table.

<u>Expected number of earthquakes</u>

In order to deal with highly active situations where the expected probability of an earthquake would be 100%, with many forecasts, the table included an 'expected number of earthquakes' figure.

- **Issues with the concept of 'less than 1' earthquake being forecasted**

When fewer than one earthquake was forecasted, rather than communicating the expected number of earthquakes as a decimal, we chose simply to write the phrase "Less than 1". This decision was taken in anticipation of confusion over what a fraction of an earthquake means (what is 0.4 of an earthquake?). For one public participant, the phrase still resulted in confusion over how there could be a number of values between 0 and 1, yet an earthquake either happens or it doesn't. They suggested using the phrase "maximum of one" instead.

Participant in public focus group 2: "[An earthquake] is either there or not there. I would put "maximum of one"

Some of the public participants also suggested that the 'less than 1' phrasing may cause some people to think everything is safe and there is no need to worry (although this discounting may of course apply to decimal expressions of the number of earthquakes too). This was a concern that was shared by some of the seismologists as well.

Participant in public focus group 1: "In both the cases the number of earthquakes is less than 1, so what would the population feel? In this case they would think that there would be nothing to expect."

Since the forecasted rate of earthquakes will be less than 1 even for high likelihood forecasts, this kind of discounting could be dangerous if it falsely reassures users during seismic sequences where the risk is high (even if there isn't more than one earthquake forecast).

Participant in public focus group 4: "I would be worried if I saw a chance equal to 39.5%. It would make me anxious. At least if I read "less than 1 in a week", the likelihood is low. I would feel calmer if I saw "less than 1" rather than 39.5."

Some participants also expressed dissatisfaction with the obscurity of the phrase "less than one", saying that the possible range of numbers that "less than one" could encompass is large and the phrasing obscures any differences between the two intensity thresholds in terms of the number forecasted. They would rather know the precise number.

Participant in public focus group 1: "When I see at the "less than 1" I see that it is the same for intensity IV or above and intensity VI or above, but I think that in reality it is different. Namely, it is true that it is between 0 and 1 but how much do they differ...I would like to see the data, I can imagine it but I would like to see it precisely."

Given these comments, this column may be less than helpful during periods of quiescence, when the rates will always be less than one, and indeed this was suggested by one of the seismologist participants. Dropping this information would not only solve the problems of confusion about this expression, but it would also reduce the density of numbers on the dashboard, simplifying the display. The rate however, should be included during seismically active periods when more than one earthquake is expected, especially since the effects of a sequence of earthquakes can be multiplicative and thus an elevated level of response might be required as the sequence progresses.

Chance of one or more earthquakes

- **Issues with the phrase ‘1 or more earthquakes’**

Public participant responses to the information communicating the chance of one or more earthquakes included similar critiques regarding obscurity caused by rounding (we used a similar style of rounding when communicating forecast percentages <0.1%), and by the phrasing itself of “Chance of 1 or more earthquakes”. Our intention with the latter was to convey to participants that multiple earthquakes are possible, however some participants commented that this “or more” is vague and does not provide any useful information since it doesn’t give an actual count of how many there will be. Other ways of conveying the idea that the forecast is not necessarily restricted to just one earthquake should be explored - this information is important to include given the occurrence of one earthquake increases the likelihood of another.

Comparison with an ‘average week’

The intention behind including a comparison to a ’normal’ or ‘average’ week was to provide context to the (generally low) absolute risk in the form of a relative risk, showing how the current forecast likelihood compared to what was “typical” for the area. We wanted to explore whether participants understood the comparison, and whether they found it helpful.

- **Issues understanding what was meant by ‘average week’**

The choice of an ‘average’ week was really made in order to provide an example of a typical, everyday risk to give context to the current forecast likelihood - whether that be higher than normal or not. Some of the seismologist participants agreed that this could be a useful feature. However, for some public participants, the meaning of the “average week” was not clear.

Participant in public focus group 3: “What do you mean by “average week”, how is it calculated? Can it change?”

For the dashboard mock-up representing a quiescent period (i.e. a time where the forecast was essentially average), further confusion was created by the fact that the table stated that there was “no difference” between the forecast and the average, with some participants wondering what exactly the table was supposed to be telling them in such instances, perhaps thinking that it wasn’t showing them anything new.

Participant in public focus group 1: “So if there is no difference, what further information is the table giving us?”

This may be in part due to the actual wording choice of “No difference”, and in part due to the lack of explanation about the context that this information was intended to provide to the user (i.e. that the likelihood is not elevated compared to a “normal” week).

One of the difficulties in explaining this is in trying to explain both the gist of this comparator (i.e. ‘normal’, ‘typical’ or ‘everyday’ level of risk) whilst also not raising questions in the audiences’ minds as to what the exact definition of that is. We had hoped that ‘average’ combined both gist and enough detail, but for some people it was not enough of an explanation of detail and for many it didn’t seem to give them the idea of its purpose as a comparator.

- **Issues understanding what 'no difference' as a relative risk means**

Another difficulty is in the case where the current forecast is no different from the average week, trying to make this information salient at the same time as not causing users to think the comparison information is meaningless and/or discount the risk entirely (since earthquakes can occur at any place, at any time and without any warning, even where the forecast likelihood is low). This is a particularly important problem to solve if such a comparison to the average week is to be retained within the dashboard, since most periods of time will indeed be 'average' and thus this lack of difference between the current forecast and the average week will be commonplace.

It is possible that a statement explaining the meaning of the "average week" and/or that earthquakes can indeed come completely out of the blue might help, although the value of adding this information must again be traded off with the fact that the dashboard is already very visually cluttered and text explanations are rarely read and understood. A better solution might be to make it very clear what the purpose of the comparison is so that the audience don't question the details so much, having understood the gist.

This situation again highlights the difficulties of creating a succinct and intuitive communication system for a subject unfamiliar to the audience, where there are many aspects requiring explanation and few existing verbal or graphical 'shortcuts' with which people are already familiar.

It should be noted that there was one public participant who did correctly interpret and take meaning from the average week comparison, suggesting that if communicated correctly, this information could well be worth retaining in future communications.

Participant in public focus group 1: "I saw the chance within the next 7 days and in an average week, and it means that everything remains normal."

- **Issues understanding the relative risks for two different intensities being the same**

It wasn't just the quiescent period dashboard for which the average week column created confusion - for the version of the dashboard showing a forecast likelihood of 39.5% there were also issues. Here, the forecast likelihood was 100 times higher than the average week for both thresholds of forecast (intensity IV+ and intensity VI+), and this description was thus written in the average week column. The primary confusion here was the fact that the likelihood for both intensity thresholds was 100 times higher than the average week (a *relative* risk), from which some public participants in focus group 3 drew the incorrect conclusion that there was in turn no difference in the *absolute* likelihood forecast between the two thresholds.

Participant in public focus group 3: "I got stuck on "chance compared to an average week: 100 times higher" and I swear I didn't understand what the meaning was, because "100 times higher" was written in both... One needs to think about it and why it is the same for both intensity IV and above and intensity VI or above…. According to me if there are no significant changes between intensity IV or above and VI and above, like in this case because in both the case is 100 times higher and "no difference", I don't know how it can help people to distinguish between the earthquakes. I don't know how it could be useful to understand the intensity of the earthquake."

This confusion may in part be due to the orientation of the table, where the eye might be drawn to read down each column more than it is across each row. As such, the information “100 times higher” (and indeed “No difference” in the quiescent dashboard) may be associated together, rather than each being associated with their respective absolute risk in the previously adjacent column. Indeed, a participant who saw the table in a different orientation, where each column contained all the information (absolute and relative risk) pertaining to each intensity threshold did not experience the same confusion. However, it is also likely in part due to lack of a clear explanation of what these comparators mean and how they can lend context to users’ understanding of the information, relating to the earlier critiques that the purpose of the dashboard and each individual component needs to be better explained to the users. It is likely that similar misperceptions will also apply to the other relative risk given in the table: comparisons with the week of the Norcia earthquake.

- **Issues understanding the direction of the relative risk**

A final misinterpretation about this section of the table that is important to mention here is that on reading the information “Chance compared to average week: 100 times higher”, one public participant in focus group 3 interpreted this to mean that the forecast was for a “good” week i.e. that the chance in the average week was 100 times higher than the forecast week rather than the opposite way around. This misinterpretation is understandable, but could lead to a very serious misjudgement of risk.

Careful design of table orientation, wording, giving just one magnitude/intensity threshold of forecast (as suggested earlier) and explaining better the purpose and usefulness of the information in the table may all aid table comprehension overall.

- **Issues about the degree to which numbers should be rounded /communicating uncertainty**

The seismologist participants also had discussions regarding the comparisons between the forecast and both the average week and the forecast in the week following the Norcia earthquake, often related to the specific wording used. One seismologist felt that the expression “100 times higher” was too vague and that it would cause them to doubt the reliability of the information, although another felt that a rounded number like this would be useful for public users. Yet another expressed that “100 times higher” implied too much certainty about the comparison, and suggested changing this to “about 100 times higher” to better convey the inherent uncertainty. It is worth noting however that research on people’s perceptions of the word “estimated” as an expression of uncertainty in other domains has shown that it has no effect on people’s perceptions of how uncertain a statement is (van der Bles, van der Linden, Freeman, & Spiegelhalter, 2020), and thus a similar reaction may be seen with the hedge word “about”.

Relatedly, seismologists in both focus groups felt that the phrase “no difference” also implied too much certainty, and could perhaps be rephrased to state “minimal differences”. This was noted particularly in the comparison to the forecast in the week following the Norcia earthquake, where it was felt by one seismologist participant that stating there was “no difference” between the current forecast and the Norcia forecast might imply that an earthquake identical to Norcia would occur, which is not only incorrect but might also result in a loss of trust when this does not happen.

Participant in seismologist focus group 1: "Saying "no difference" might be too categorical...maybe say "minimal differences?". The same applies to "100 times higher", maybe this could be written as "about 100 times higher" to stress the uncertainty of the information."

Comparison with the likelihood of an earthquake 'in the week of the Norcia earthquake'

In order to help put the current forecast in context, the table included not only a comparison of the forecast likelihood to the average week in the location in question (to give an idea of a 'low' or 'typical' risk), but also a comparison to the forecast in the week following the 2016 M6.2 earthquake in Norcia, Italy, which would be familiar to the Italian audience (to give an idea of what a 'high' risk might be). We wanted to understand firstly whether this information was comprehensible to participants, and secondly whether the Norcia earthquake was a salient and useful comparator, or whether another earthquake (or another domain entirely) might be thought of as being more useful. Feedback regarding this comparator information was collected from both public and seismologist participants.

- **Norcia earthquake as a choice of comparator**

Some of the public and seismologist participants felt that Norcia was a good comparator, since it was a fairly recent earthquake that may thus have resonance for both old and young generations. One participant commented that it had more salience to them as a comparison than the comparison with the average week since it was a real, concrete event:

Participant in public focus group 2: "I found the comparison with the Norcia earthquake especially useful because it does not talk about "Average week" but of a specific earthquake that happened in a specific interval, and you tell me there is "no difference", so ok, now I am worried."

However, thinking from a more technical perspective, one seismologist participant had concerns that Norcia was an unusual case in being the third in a sequence of larger earthquakes, and thus might not be "typical enough" to provide a useful frame of reference.

Another public participant discussed how particular earthquakes, even if not experienced, are still talked about and thus can still have salience even if they were not experienced directly.

Participant in public focus group 3: "All have a historical memory of an earthquake experienced or spoken of… For me, we talk about the 1908 earthquake that hit Reggio Calabria and Messina, that then caused a tsunami, and it was devastating and even now we talk about it. It is not rare that we mention it when there are narrow streets with a building in the middle, full of people, we'd say "If there was an earthquake like 1908, no one would survive."

This is encouraging for a country like Switzerland, where there have been very few large earthquakes in recent years, but where there are some in the historical record. Indeed, one of the seismologist participants suggested the Basel earthquake of 1365 as a possible comparator for Switzerland, although further research would be needed to assess whether this was a salient comparator for the Swiss population - and what the calculated forecast likelihood of another event might have been in the week after it - before it was adopted.

Some public participants suggested that the comparator earthquake could change depending on the location of the user, thus showing a large and/or the most recent local earthquake that that individual might have actually experienced, rather than simply heard about. Other possible Italian comparator events suggested in another public focus group include the 2009 L'Aquila earthquake (M6.3), the 1980 Irpinia earthquake (M6.9), and the 1976 Friuli earthquake (M6.5) - although again, for the earlier events, the forecast likelihood would have to be reconstructed.

- **Issues with participants confusing likelihood and intensity**

It is important to note that there was a comment from a public participant that indicated they were conflating the comparison of the forecast *likelihood* just after the Norcia earthquake with a comparison of forecast *intensity*.

Participant in public focus group 4: "If the chance is 100 times lower, I would not be worried about it. It is like hearing a lorry pass by the front door, a slight shaking, so I would not be worried about it."

Whilst we anticipated this misunderstanding with the map component of the dashboard and thus added an explanatory statement to the map to make things clearer (see later), we did not anticipate it regarding the comparator information in the table. In retrospect, this confusion was possible to anticipate, given that 'risk' is a feeling fed by the combination of likelihood, impact and vulnerability - and mentally separating those components is unfamiliar for most people.

Future designs should thus make it clear that the information being depicted is the likelihood (which can vary) for a set threshold of earthquake intensities (or magnitudes) and try to communicate that.

- **Issues with understanding that the probability in the week of Norcia was a forecast (in the past), for another earthquake after the Norcia event**

One thing to note more generally about the Norcia comparison information in the table was that although we had chosen to use a forecast from the week immediately *after* the Norcia earthquake (when the earthquake likelihood forecast would be at its highest level, as an example of a 'high' risk), this information was not made clear to participants. Since the concept of a forecast likelihood being higher *after* a major earthquake is not immediately straightforward to understand, we thought it might be best to keep the description of the Norcia earthquake forecast vague in timing, thinking that the precise timing of the forecast in relation to the earthquake might not matter to participants. This decision turned out to be misguided however. Indeed, some participants (both public and seismologist) asked when specifically the forecast was from.

Participant in seismologist focus group 2: "When you say the week of Norcia earthquakes, do you mean the week starting from the time of occurrence of Norcia for the next 7 days or do you place the Norcia earthquakes within this week?"

This could not only lead to the misperception that these large events can be reliably predicted, but also could provoke anger that if such a forecast was known about prior to the Norcia event, why were people not warned or evacuated, as expressed by one of the public participants:

Participant in public focus group 1: "I was thinking that 40% is very high. If this was known before [the Norcia earthquake], it would have been possible to intervene or secure some people or warn the population. What I am wondering is whether this data was calculated later on, or if it can be estimated before the event. So with the range of a week, can one give the percentage chance?"

In turn, there were parts of the focus group discourse that indicated some participants assumed the information was the forecast of the Norcia earthquake itself, rather than that produced in the wake of it. This lead to confusion for one participant when they looked between the forecasted number of earthquakes for the present forecast location ("Less than one"), and how this compared to Norcia ("No difference"), since Norcia was an earthquake that actually occurred and thus - in their mind - should not have happened in a week labelled "Less than one". Interestingly, this relates to the aforementioned concern of the seismologist participants regarding the definitiveness of the statement "No difference", and one in particular who suggested this might incorrectly imply an earthquake identical to Norcia is forecast to occur (see above), and also the problems that people have with probabilistic thinking.

2) Text Explanation

The text explanation component of the dashboard (see Figure 9) was designed to carry the same information as the table, turn the key concepts from the table into a full sentence, then provide an alternative presentation format that translated the probabilities into expected frequencies (as previously mentioned, a common method of making probabilities more understandable), which included positive as well as negative framing (the chance that there would NOT be an earthquake). We were particularly keen to have feedback on whether expected frequencies were useful in this context given the problems (discussed previously) of explaining the reference class (possible ways that the next 7 days could turn out, which we simplified to 'weeks like this').

Figure 9: The 'text explanation' format within the dashboard

- **Issues with repetition of the same information in sentence form**

Although some participants found certain aspects of the explanation a useful complement to their understanding of the table, several participants in the public and seismologist focus groups wondered what the point of the explanation was, and what else it showed in addition to the table itself. One seismologist participant suggested that the fact that the text explanation shows the same information as the table but just in a slightly different way may cause confusion, as one is not sure if what they are seeing is indeed the same information or if they have just misunderstood something.

Participant in seismologist focus group 2: "You read this and then you read the explanation and wait, this is the same that I have read. Is this a duplicate information? Is it something new? So to me it is a little confusing."

Some public participants found parts of the text information obvious, for example where the likelihoods were restated in more explicit terms: "The chance of one or more earthquakes of intensity IV or above is 39.5%". Some even considered this information to be almost insulting, indicating to them that we felt they would not understand the numbers. The second seismologist focus group suggested dropping the percentage format and relative risk information and only giving the two sentences translating to expected frequencies and alternative framing. Public and seismologist participants also agreed that giving people both the table and text information increased the potential for information overload, without gaining much benefit from the alternative presentation (whilst other components of the dashboard were giving additional information).

- **Expected frequencies format highlighting issues with probabilistic thinking**

Some participants in public focus group 2 found that the expected frequencies format highlighted the fact that the dashboard was showing probabilistic forecasts, and that they didn't understand that. These participants again seemed to interpret the forecast as being a prediction of the specific timing of an earthquake, and were thus dissatisfied that the expected frequencies made it clear that there was no means of knowing the precise day on which the earthquake would occur:

Participants in public focus group 2:

"I find the explanation on the bottom, besides useless, also nerve-wracking. In which of these 395 weeks could this earthquake occur, do you understand? It doesn't contextualise anything and it leads me to a condition of uncertainty. It does not give me certainty, some landmarks."

"What xx said is true, among the 395 weeks, in which one will it happen? I have to stay alert for all the 1000 because I don't know when it is going to happen. It is useless."

Clearly, these participants were correctly interpreting the information being given to them - and this format may have made it much more explicitly clear than the probability statements. What they were reacting against was the gap between the (probabilistic) information that they were being given and the (definitive) information they expected (or wanted).

- **The usefulness of the expected frequencies format**

One seismologist suggested that it was difficult to conceive of 1000 weeks as a reference class in the expected frequencies format ("a 39.5% chance means that out of every 1000 weeks like this we'd expect an earthquake to happen in 395 of them and no earthquake to happen in 605 of them"), and that the format was therefore not very user friendly.

3) Map

The map component of the dashboard (see Figure 10) aimed to provide geographic context for the forecast at the selected location, showing how the likelihood of an earthquake of intensity IV and above varies across the overall country.

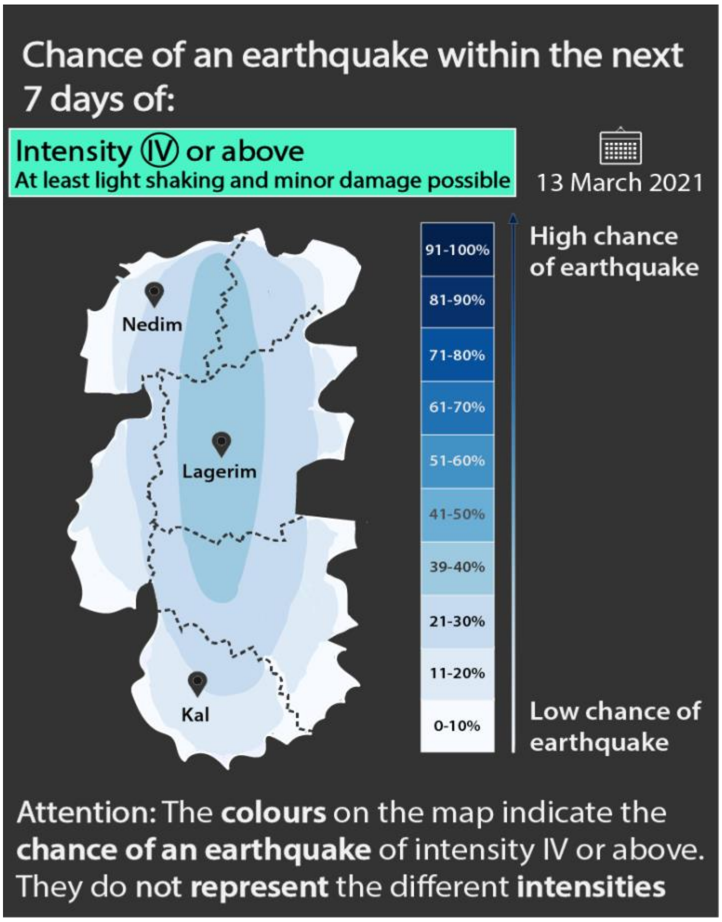


Figure 10: The 'Map' component of the dashboard

- **Confusion over whether the map shows likelihood or intensity**

We anticipated that people may incorrectly assume the map was showing variation in intensity rather than likelihood (because the 'contours' appeared to be spreading out from an epicentre and the most familiar context of this sort of presentation will be in showing the effects of an earthquake event). We had therefore included a sentence underneath the map to attempt to avoid such misunderstanding: *"Attention: the colours of the map indicate the chance of an earthquake of intensity IV or above. They do not represent the different intensities."*

Some of the seismologists felt that the map was clear and provided a simple and intuitive first level of information, and some public participants noted specifically that it was clear to them that the map was communicating likelihood and not intensity. For others however, there was still confusion regarding which the map was showing, perhaps because they did not read this explanation and/or because the green banner at the top of the map detailing the intensity threshold was more visually eye catching than the title detailing that the map depicted the chance of an earthquake within the next 7 days (as suggested by participants in the second public focus group).

It is likely that the blue colour scheme chosen for the map also added to the confusion however, as it corresponded with the colours of the lower intensities on our visual Mercalli scale (I, II and III). Such colour choice may also have led to confusion about which threshold of intensity the map is depicting too; when asked this question, some participants answered that it showed intensity II or III (the colour blue on the Mercalli scale). While some participants did ultimately notice their mistake, others kept with the assumption that the map was displaying variation in intensity across the country.

Participant in public focus group 2: "There is the possibility, that you have evidently foreseen, of confusing the colours on the map with the colours of intensity, because the first thing I thought was that [it showed intensity], and then I went and read "chance of earthquakes" and I said ok, the map is based on that, but this is a problem as people do not always read sentences, or do not focus on the little words."

Ultimately trying to use a map to represent something different from what the audience are familiar with (the effects of an earthquake) when the patterns are likely to look similar, may be destined to cause confusion and misunderstanding.

- **Issues with legibility of the colours**

Further comments regarding colour pertained to the level of saturation, particularly for the quiescent period dashboard, where some public participants found it difficult to distinguish between the different colours and thought perhaps that the map was unfinished. They also struggled to match the colours on the map with the colours on the associated scale, which can be an issue when too many categories of colour are used in a visualisation. Some of the seismologists in the first focus group also commented that the map colours could be made more saturated.

The problem is really trying to illustrate such potentially variable probabilities on a linear scale. Because of the known difficulties audiences have with interpreting logarithmic scales (discussed previously), and the fact that small differences in absolute probabilities are not necessarily important to highlight (as they may be too small to be relevant), we were keen to try to use a linear one. This meant that we were left with a scale which covered the whole range of saturation but where most maps (which would usually be showing low absolute probabilities, during quiescent times) would be using only the lowest couple of categories, with extremely pale colours.

- **Issues interpreting the locations on the map**

Some issues arose that may have been mainly due to the fictional location we created for testing. While we had split the map into regions, divided by dashed lines, we did not make it clear that the fictional location of Lagerim, to which the forecast pertained, was a city and not the name of a whole region on the map. Participants also wondered why only three of the six segments of the map contained a name (Lagerim, Kal and Nedim), likely also due to the misperception that these names referred to regions and not to cities within them, as we had intended (and hence marked with a location dropmarker). There were requests to label each region or province along with their capitals. This might make the map cluttered, and may not be necessary when the map represents a familiar country.

- **Map interactivity**

Some public participants suggested that the map could be made interactive, such that users can zoom in and see how the forecast changes between locations. In a similar vein, one of the seismologist participants suggested making the map dynamic, such that a location could be selected alongside a chosen radius around it, and the forecast presented for that area (again, like https://www.richterx.com/?go=forecast), although another expressed concern about the ability of a website server being able to handle this type of interactivity in a timely manner.

Whilst the idea of interactivity and customisability is always initially attractive, as previously discussed, it can in fact increase confusion and complexity. In particular, since there are several parameters that can change an OEF forecast probability quite dramatically (location, intensity/magnitude threshold, time frame considered, area of forecast unit), allowing customisation of any of these may cause audience confusion about why the numbers are changing so much. As already noted, parameters such as time frame and forecast area, where a longer time frame or larger area will increase the forecast probability simply because there are more possible opportunities in space and time for an earthquake to occur, can increase probabilities to very high numbers - a concept that may be difficult for users to make sense of 'at a glance' and which are not necessarily helpful variables for most use cases. Instead, we want to focus the audience's mind on the temporal changes in probability at a fixed location, area, threshold and timeframe.

Interestingly, the discussion about interactivity in public focus group one revealed that participants were imagining using the dashboard for different purposes, and at different times in the earthquake cycle. One participant mentioned that they would like the default view of the map to be set to the entire country, such that after an earthquake, they could search the affected area and get more information; it was not clear whether they were seeking forecast information (likely to be a raised likelihood of further activity) or information about the impacts of the event that had happened (in which case the purpose of the dashboard would have been misunderstood by this individual).

Participant in public focus group 1: "I would view the website when I heard of earthquakes happening in Italy, so I would always prefer the whole of Italy [to be the default], then I'd go and select the area."

Another was more explicitly imagining using the dashboard as a tool for forecast in their particular area (we had mentioned in the introduction to the tool that it was similar to a weather forecast), and so wanted the default view of the map to be the location of the user:

Participant in public focus group 1: "Eh, it depends, because you were mentioning the weather forecast before and I usually look at my location."

- **Choice of forecast area parameters**

One of the seismologist participants pointed out that the map needed to reflect the grid of areas over which the forecast probabilities were calculated (e.g. 10km squares). Showing the probability contours as smooth curves did not accurately reflect the underlying model. However, the difficulty of how a forecast should be interpreted when the location in question is on the border of two units of different colour needs to also be considered, something that has been raised as an issue in natural hazard forecast maps (Becker et al., 2019). Indeed, one public participant appeared to interpolate probabilities near boundaries, which may or may not be a justifiable interpretation of the data.

Participant in public focus group 3: "Kal, it is between 10% and 20%, maybe 19%"

Increasing the size of the grid area over which the forecast probabilities are calculated would raise the absolute probabilities of any seismic activity, and is probably the best parameter to alter in order to make the probabilities more easily manageable, since the geographical resolution of such forecasts does not need to be particularly small (if communicating magnitude, rather than intensity), whilst the week-long time frame, and the threshold of earthquake being considered are more important variables to get right for the audience.

- **Choice of top end of scale**

One of the public participants suggested that perhaps the map scale (which went from 0-100%) could be truncated, since the probability in any location is unlikely to be 100% (at least during periods of quiescence). They were inspired by the fact that there was a statement underneath the risk ladder (see below) stating that the chance of an earthquake Intensity IV+ would be unlikely to be higher than 25% in an average week in any location (and indeed the scale on the risk ladder was cut off at 25% when illustrating a quiescent probability).

Participant in public focus group 3: "Since a probability of 25% is rare, I wouldn't [scale the map to] 100%. Maybe to 40% or 60%, I don't know. This is to show that, at 39.5%, we are high in likelihood. Besides that, according to me you will never be sure there will be a 100% likelihood. If you already know that some intervals are not reachable, it does not make sense to put them."

The addition of this definition, and accompanying truncation of the risk ladder, was done in order to be able to 'zoom in' to illustrate more clearly markers in the lower portion of the ladder. The colours remained the same as for the 0-100% scale. Truncating the scale for the map would not serve the same purpose - of showing greater resolution - and it would also not change the colour saturation (so no potential to allow greater discrimination). The participant's instinct that simply putting a lower value as the 'top' of the visual scale might cue that these were 'high' values may indeed be correct, as similar effects have been found in risk ladder research (Sandman, Weinstein, & Miller, 1994a).

This comment also reveals a potential misunderstanding from the risk ladder, which is discussed in more detail in that section, below.

The other point raised by this same participant is also worth noting; the fact that the map scale runs to 100% is that this conveys that forecasts of seismic activity can sometimes reach absolute certainty (100%), which is never the case - although in some earthquake sequences the probability of an aftershock does approach 100% closely.

4) Risk ladder ('Thermometer')

As already mentioned, one of the key challenges of risk communication is providing appropriate and balanced context to allow people to assess how they feel about the risk. Such context can be provided by comparing the likelihood component of the risk with other likelihoods of the same event - at other times (such as the relative risks presented in the table format, e.g. comparing the forecast probability in the given location with the average probability in that location), or in other places. Based on previous work done on choosing appropriate comparators for COVID-19 risks (Freeman et al 2021), we decided to test comparator locations chosen to represent what people might perceive as 'high', 'medium' and 'low' risk cities for seismic activity, and present them in the format of a risk ladder, a well-known format for presenting comparator risks (I.M. Lipkus & Hollands, 1999; Stallings & Paling, 2001). See Figure 11. Our aim was to allow people to gauge the 'feel' of the riskiness of the forecast for the week ahead in their chosen location instinctively, by comparison with the familiar 'feel' of the normal level of risk in the comparator cities.

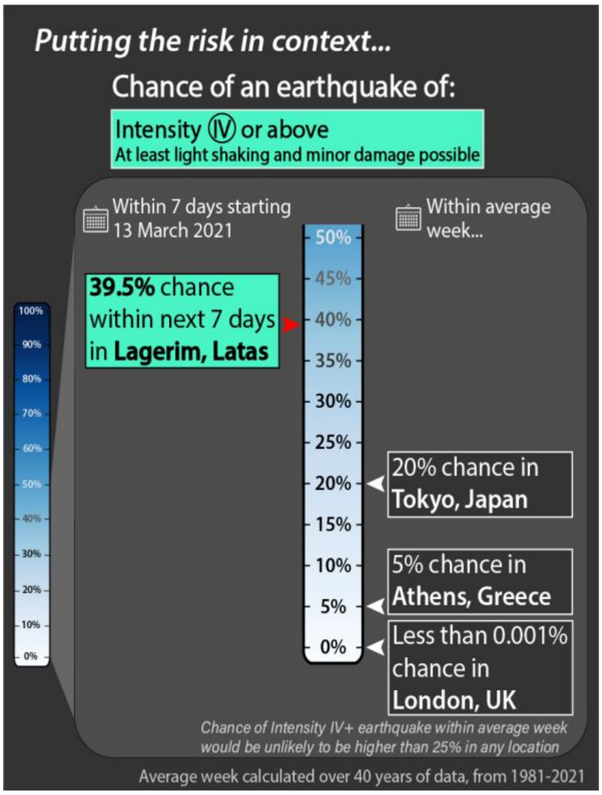


Figure 11: The risk ladder ('thermometer') component of the dashboard

- **Issues understanding, or understanding the purpose of, the risk ladder**

Some public participants expressed confusion about the risk ladder image.

Participant in public focus group 2: "I don't know, I don't understand it."

Similarly, some of the seismologist participants found it harder to understand than other sections of the dashboard.

Participant in seismologist focus group 2: "I think this is the least clear [component of] the whole dashboard because for example I didn't understand the percentage for Tokyo, Athens and London referred to the next 7 days, so I think, which time period. I have to read everything to understand what you want to communicate with this one. The other parts took me less time to understand the general meaning, this one in my opinion is the least clear."

In the case of the seismologist participants in particular, this perceived lack of clarity could be because of the comparative unfamiliarity of the risk ladder format. All the other formats (text, tables and bar charts) are extremely familiar to all seismologists (and many members of the public) whilst the risk ladder would not be instantly familiar and so likely to result in a subjective feeling that it was not instantly understandable. However, some of the perceived lack of clarity is also likely to be due to lack of clarity of the purpose of the format.

The potential disparity between subjective and objective comprehension of the format was illustrated by one public participant who expressed that they did not understand the image, but actually did appear to have a good sense of what the risk ladder was attempting to communicate:

Participant in public focus group 2: "I found it very difficult to understand. I'll tell you what I understand but I am not sure I understood properly. This is a contextualisation. You see the 40% probability and then you see Tokyo has 20% to make you understand Tokyo is enough hit, so if Tokyo has 20% and you 40%, let's say that you should worry. I think it is not priority information but it is important because it helps you understand whether to be worried or not. Even if it does not make sense because I am speaking as a person who hasn't studied statistics and who has nothing to do with this stuff."

Participant in public focus group 2: "In the common imagination people think of Japan as a high seismic risk city. Reading this information you see Tokyo, connect it with earthquakes and from there, you start your thinking."

Participant in public focus group 3: "Let's talk about hurricanes instead of earthquakes. If my city had a higher likelihood than New Orleans, I would be worried because in my perception, New Orleans is at high hurricane risk. Tokyo is the most famous city, but Japan in general is famous for its seismic events so the fact that [Lagerim] has a percentage almost double that of Tokyo… For me it is a relevant information. With regards to the [25%] subtitle underneath, it stresses that we find ourselves in an unlikely situation. It contextualises better the data. Having some examples of high seismic risk places, but also having some with low seismic risk can be useful. I have never personally connected London with a place at high seismic risk, so the fact that it is there, as an example of "not seismicity", for me it is also fine. The part that is *written below, it does impress me, namely Tokyo is well known as an area at high seismic risk but the fact that Japan is under 25%, we are above 35%, it worries me."*

Participant in public focus group 2: "it is clear the aim, namely when you tell me the cities, the aim is clear, to contextualise, Tokyo is the example to which you should be drawn and London is the normality... as a ignorant person seeing that in Tokyo there is a 20% chance and I have 40%, it gives me that sense of alert, it helped me to contextualise the number so I think that the visualization was successful"

This is encouraging and suggests that the risk ladder is, to some extent, serving its purpose for some participants, and that improving the design of the graphic and clarifying its purpose and uses for the audience might help build confidence in users who may be unsure of their interpretation.

- Choice of comparators

Again, the broader purpose of the communication - to allow them to calibrate their own feeling of the risk in their location in the coming week by comparing its magnitude with that in other cities chosen to be those we anticipated they might have an 'innate' feel for the level of risk - was not clear to some participants. They therefore questioned why information about other cities was included.

Participant in public focus group 3: "Why put other cities? Why should I be interested in other cities and the percentage that there is in Tokyo if I am here? I have no idea. But if I will travel, it is something nice to know and interesting."

Participant in public focus group 3: "I wouldn't put the comparison, namely I know that there is a likelihood [in Lagerim] of the 39.5% and that's all and this is ok. The likelihoods in Tokyo, Athens and London, it doesn't give me anything more than what the explanation does."

For others participants however, the purpose was clearer, particularly for the higher risk comparator of Japan, and did seem to be doing the job of helping people make meaning of how high or low the risk was. *S*ome of the lower seismicity areas, however, were less deemed less helpful to them. Part of the problem with choosing comparators with low seismicity is that they are inherently less salient in the seismic context - although their position on the risk ladder gives the audience the information they need.

Participant in public focus group 1: "Honestly I didn't know London was at low seismic risk but I am happy for them.

Participant in public focus group 3: "Maybe I am ignorant, maybe there was a memorable earthquake in Athens or in London. I know about Tokyo, yes, but, I don't know. Besides Tokyo, that we know is a seismic area, the others, I don't know."

Participant in public focus group 4: "I did know about Tokyo and I think all are aware of the terrible earthquake Tokyo went through and also Athens, and Turkey have always been hit by earthquakes. In London I did not know that the possibility of an earthquake was 0. I think that in a next life, I will be thinking of move to London if it is so safe"

Indeed this concern about the low salience of some of the lower seismicity comparators was shared by one of the seismologists, who anticipated a slightly different use for comparators than perhaps the risk ladder is designed to give.

Participant in seismologist focus group 1: "Comparing the likelihood of your location with others is smart, assuming that the end users know the seismic probability in London. In the thermometer I think our aim is to make people think of their likelihood and make it in comparison just to make them more aware… and I do not think that the comparison with other earthquakes communicate the "insight", wow, I did not think of it, that we want to give rise to."

Some participants, both public and seismologists, questioned why the cities shown were chosen and suggested it might be nice for users to be able to choose which were displayed.

Participant in public focus group 1: Maybe if it was something interactive, it would be nice to write down the city and get a %."

Participant in public focus group 2: "Ok, it tells me Tokyo, it tells me Athens, London. From the point of view of a possible application it would be nice to choose what you want to insert. The capitals are there as default options and then you can choose what to insert. But the fact that in Athens there is the 5% and then I am asking you and then in Paris? In Los Angeles?"

Participant in seismologist focus group 2: "Giving a context can be a good idea, but giving these comparisons, no. Also because why Tokyo and Athens and not Rome, something different."

Participant in seismologist focus group 2: "I was a little surprised by London... Why not indicate a city, for example Switzerland, where the hazard is less than Athens. In London I don't know, in Switzerland, in France, these kind of risk regions [could be better]. London is a little surprising."

Relatedly, some of the public participants mentioned they would prefer comparators within their own country, and some stated that they were not aware of earthquakes outside of their home country of Italy:

Participant in public focus group 2: "I understand the reason why there is Tokyo but London and Greece, if the website will be in Italian, rather than Athens I would have preferred L'Aquila, or the Basilicata or wherever there have been important earthquakes throughout history. About Greece that is 5%, rather I would have put either where there are or where there aren't."

Participant in public focus group 4: "Out of Italy I didn't know, I can think of earthquakes in Italy."

Participant in public focus group 4: "Italian cities are the ones we are more familiar with. We remember them more. But it is ok to put something like Tokyo because it gives you a further comparator."

Future work should help identify a range of potential comparators which would contain enough that would be salient to a broad audience in each country (ensuring that it contains the fullest spectrum of likelihoods possible to serve its purpose)

One of the seismologists noted that whilst they saw the value of the comparators, in some instances where the forecast percentage was comparatively low, they were concerned that it could mislead people into thinking there is no risk, when in fact earthquakes can occur at time, without warning, even where forecast percentages are low.

Participant in seismologist focus group 2: The comparison can be helpful, it depends, if the [forecasted] risk is much higher than in Tokyo then it is helpful to know what the risk is in Tokyo, kind of comparison, but if the chance is lower, maybe at the bottom end, then this information could be misleading, [making one] think that there is no risk, even though there is."

It was also suggested by one seismologist, as already mentioned, that cross-domain risk comparators could be used in addition to within-risk (earthquake risk) comparators.

"Show two thermometers, one that compares the likelihood with other likelihoods of different earthquakes and another comparing with the likelihoods of other negative events."

And on a technical note, one seismologist warned that it was important to ensure that the city comparator risks were calculated over the same grid area size as the forecast.

Participant in seismologist focus group 2: "I don't think this could be helpful also for the spatial extent because on one hand you have regions, the other side you have the city. I don't get the real extent of the cities."

- **Issues understanding the concept of the 'average week' in the comparator cities versus the 'coming week' for the specific location in the forecast**

It appears that some participants thought that the risk ladder was showing the forecast for the current location on the left hand side, and the forecast for other comparator cities on the right hand side, rather than the average week for each of the comparators on the right hand side. This was true of both public and seismologist participants.

Participant in seismologist focus group 2: "For me as well, at first it was not clear whether it was within this week or within an average week but then again I saw it, but within an average week was one of the last things I saw."

It's not clear how important this misunderstanding is, since the majority of the time the current forecast in the comparator cities will equal the average forecast, but the purpose of using the average week in comparator locations was in part that this would provide a steady baseline to which any varying forecast could be compared, rather than having the comparators themselves vary depending on the seismic activity in the comparator locations, which would not only add many more numbers for users to contend with, but would also mean that the comparators would no longer provide a consistent frame of reference.

Some seismologists questioned whether such a difference in the comparators was even valid:

Participant in seismologist focus group 2: "Within the next 7 days" or "within an average week", otherwise they are two different concepts not really comparable.

Participant in seismologist focus group 2: "The two information are different, when we look at the 39% we think that is higher than in Athens but it is not exactly true so because it is true information, not the same but put in the same diagram can believe that this information is the same, but it is not true."

Participant in seismologist focus group 2: "Maybe I would also prefer to have the earthquake risk forecast for the current week in Tokyo, Athens to make the information really comparable"

Ideally, the two sides of the thermometer need to be more clearly differentiated, going beyond just simple labels and perhaps using other visual techniques such as colour and hue to highlight this difference. Another approach might be changing the wording of the description of the two sides slightly. As it stands, the forecast is described as "Within 7 days starting 13 March 2021" whilst the average week side is described as "Within the average week". The fact that the key difference between the two is the forecast vs the average might be made clearer by changing the average week label to read "Within an average 7 days". In addition to this wording change, perhaps the average week for the location in question could also be labelled on the right hand side of the thermometer during active seismic sequences, when the forecast is likely to be higher than the average (unlike during quiescent times when the forecast will be very similar to the average). This would further highlight that what they are seeing on the left hand side is not the average but a forecast, and would tie the risk ladder to the table format, which also shows the average week for the forecast location.

As with the table, there was also some confusion amongst some of the public participants over what was meant by the 'average week', despite our attempts to dispel this by including a definition at the bottom of the image. However the definition was intentionally designed be small so as not to distract from the main visual (one of the seismologist participants suggested making the definition accessible on click), so it is possible it was not seen by all participants.

- **Misunderstanding that the 'maximum' 25% likelihood only applied to an 'average' week**

One potential misunderstanding of an aspect of the risk ladder format became clear during a discussion of the map, when a participant back-referenced knowledge that they thought they had gained from the risk ladder that the highest probability of any location at any time would be unlikely to exceed 25%.

Participant in public focus group 3: "Since a probability of 25% is rare, I wouldn't [scale the map to] 100%.

This participant appears to be referencing the information in the footnote below the risk ladder which states that "Chance of intensity IV+ earthquake within average week would be unlikely to be higher than 25% in any location.", and possibly confusing the 25% upper limit for an *average* week in any location with what is possible for a forecast at any time - which could range very close to 100% during seismically active sequences. This highlights again the need for a clear definition of what the "average week" means, and a clear visual distinction between the ‘current forecast’ and ‘average week’ sections of the risk ladder (see later). One of the seismologist participants also noted that the definition does not make clear whether we are referring to the country of the forecast (Latas), or the world generally.

There were other discussions too, that indicated some level of misunderstanding of what the average week meant and how it could be used in concert with the forecast. One participant could not understand that the chance in an average week is unlikely to go above 25%, yet the forecast they were shown was for a chance of 39.5%, indicating once again the distinction between the concept of the average week statistics and the forecast is not clear.

Participant in public focus group 2: “In Lagerim the chance is equal to 40% so you are telling me that it is unlikely that the chance can go beyond 25% and then I have [a chance of] 40%. It is a little contradictory... All this to tell you that it is not intuitive, because I go and read [the definition] below and then I realise I have 40% chance.”

- **Confusion about the calendar icon**

In one public focus group some of the participants mistakenly thought that the time frame of the risk ladder was customisable, perhaps misled by the inclusion of the calendar pictogram next to the date of the forecast

Participant in public focus group 2: “Within an average week” is put there because then there could be a calendar in which you can change the time frame, e.g. 7 days, 1 day. Or is the 7 days time fixed?”

We had included the image of the calendar on several of the visualisations as we thought it would help users quickly navigate to the place where they could see the week of the forecast in question, however given this misinterpretation and our subsequent realisation that it looks very similar to many dropdown menu calendars in other visualisations and apps, it could well be worth removing it in future iterations.

- **Issues with cropping of the scale on the ladder**

One problem common to many risk ladders is that the probabilities of most interest are not evenly distributed along the scale, and so a non-linear scale would be ideal for allowing dig clear distinction of clustered risks - but potentially misleading graphically as one of the points of a graphical scale is to guide people’s perception of the magnitudes visually (Sandman et al., 1994a).

This problem is one that is true of a seismic risk ladder. During periods of quiescence (the vast majority of the time), the likelihood of an earthquake of intensity IV or above is unlikely to exceed 25% at any inhabited location or reasonable area, whilst during periods of heightened activity the probability could (in some places) approach 100%. Following previous work on communicating personal COVID-19 risks (Freeman et al., 2021) we chose to display a cropped version of risk ladder for forecasts during quiescent times, allowing clearer comparisons between probabilities clustered in the low part of the scale without distorting the linearity of the scale. In anticipation that participants might wonder why the thermometer was cropped, however, we chose to include a smaller, full (0-100%) thermometer to the left of the graphic, with a box linking it to the central thermometer to indicate that the main image was zoomed-in.

Some participants did not like the cropped thermometer however. One of the public participants did not understand its purpose for example, although it was not clear whether this was due to the design of the risk ladder as a whole (e.g. the way the zoom was depicted) or because or other more general confusion such as why there were two thermometers, or why one was truncated and one was not. One of the seismologist participants also suggested the small thermometer could be dropped from the risk ladder image, whilst a public participant suggested making the thermometer interactive such that users could zoom in and out, removing the need for the smaller reference thermometer.

- **Issues distinguishing between colour shades**

Another design suggestion, in common with the map which also used the same saturation scale to indicate the range of likelihoods from 0-100%, was that it was difficult to distinguish between the colours, and that this reduced the impact that the comparators had on user interpretation. Similar changes as suggested earlier (using multi-hue scales, increasing saturation) might solve this issue.

Participant in public focus group 1: "The zoomed part is more or less all the same colour from 0 to 20. Tokyo is 50 times more likely but for the colour we seem to be the same. It does not make me think that in Japan there is such a higher seismic risk."

- **Questions about the purpose of the communication and intended emotional effect**

During the discussion, a participant who had experienced the L'Aquila earthquake was frightened by the 40% forecast, and felt that the risk ladder format did not suit the purpose of the communication, which in their eyes should be to convey a sense of resilience. While this again stands as a misinterpretation of the purpose of the communication from our perspective, it is important to note the emotional effect that higher likelihood forecasts may have on those with traumatic past experiences of earthquakes, and consider wording changes or additions that might alleviate this to some extent. Interpreting the 'appropriate' level of emotional threat of the forecast and balancing the competing aims of 'reassurance', 'warning' and 'pure information' will be extremely difficult however, and will require careful discussion and ethical decision-making on the part of the communicators, as well as design and testing.

Participant in public focus group 2: "The communication is not efficient and effective if I, as a user, don't understand it. I have to understand it. The communication should be precise, it should not bring me to understand, it has to be precise because then each of us understands what (s)he wants. I hadn't noticed the 25% and that sentence behind it frightens people. That sentence frightens the average citizen. I found this graph really useless, no I found it dysfunctional toward a communication that should give a capacity of resilience."

- **Perceptions of manipulation and trust when information is misunderstood**

One final point highlights the importance of trust in these kinds of communications, and how communications that are perceived as unclear, or emotionally manipulative, might undermine trust. Here, a public participant was confused by the meaning of the definition and the 25% statistic, with regard to the explanation for the cropping of the risk ladder, and this led them to wonder that there may be some manipulative agenda behind the communication.

Participant in public focus group 4: "I took a lot [of time]. I kept on reading that sentence there and I kept on to try to understand why this percentage cannot overcome 25% in any location. It is strange, I couldn't understand it and I was wondering if it was something made up that have already taken for granted or there is a motivation."

Reading this quote closely suggests that it might be useful to more clearly acknowledge the uncertainty about the 25% statement. Although the word 'unlikely' was used to convey that the 25%is not an absolute limit, it might be worth stressing the fact that this is a general estimate that will not always apply. Avoiding the absolute term "any location" may also help. Such clear acknowledgement of uncertainty might enhance trust in the information, rather than undermine it.

5) Bar chart of past seismic activity

The aim of the bar chart (see Figure 12) was to provide temporal context to the forecast by showing previous recorded earthquake activity. In the dashboard representing a hypothetical forecast from a quiescent time, the chart showed the recorded number of earthquakes in the location per *year*, from 1980 onward. In the version showing a hypothetical forecast during a seismically active period, the number of earthquakes per *week* was shown, for the preceding 12 weeks.

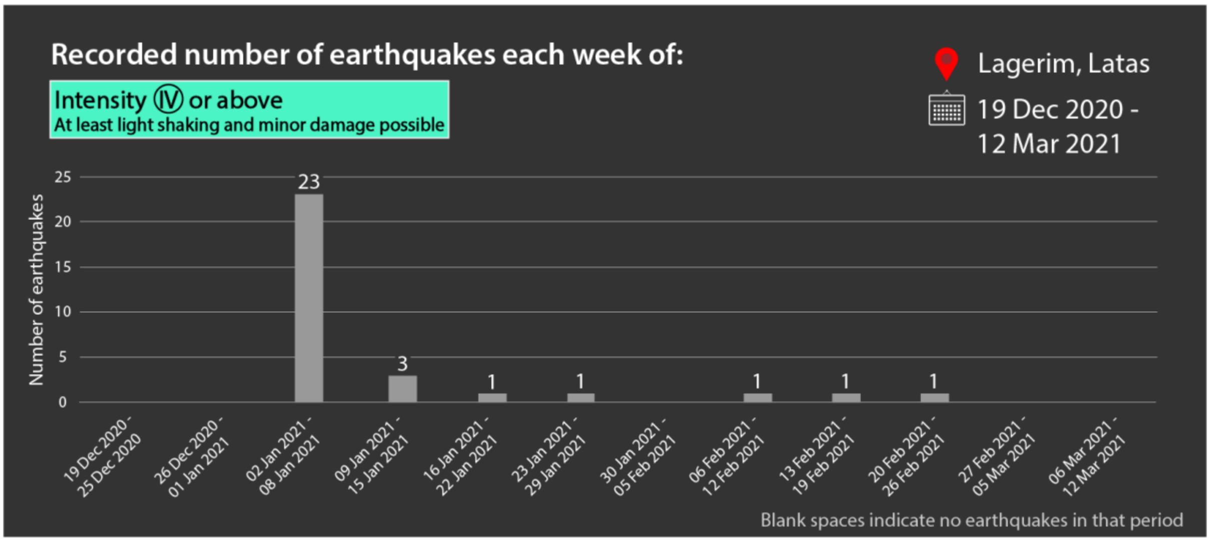


Figure 12: The 'bar chart' component of the dashboard

- **Perceived as easily understood by some, but others struggle**

Some of the seismologist and public participants liked the bar chart, commenting subjectively that it was clear and comprehensible.

Participant in seismologist focus group 2: "I think I like the chart, it is not too much information, interesting to see. I like it."

Participant in seismologist focus group 2: "It is an information that can be useful to understand the rest of the dashboard, to see right now this is the risk and this is the evolution of the earthquake"

Although one public participant seemed to be confused about whether the bar chart was showing counts of earthquakes, or variation in intensity between timepoints. This may partly be because of the visual prevalence of the green intensity banner, as discussed previously.

Participant in public focus group 4: "If it already says "intensity IV or above", how is it possible that they are reporting a scale from 0 to 90? But does it show the number of earthquakes or the intensity?"

- **Issues understanding the purpose of the chart**

Although they probably understood the information, several public and seismologist participants did not seem to interpret the bar chart as a source of additional context to the current forecast as we'd intended, and questioned its usefulness, at least as it relates to forecast information.

Participant in public focus group 2: "It is an aseptic table, it gives me a statistic of what happened between 1980 and 2021 and ok... It does not add or take away anything. If it was missing, I wouldn't miss it."

Participant from public focus group 2: "Knowing the past [earthquakes], I will not know the future ones. It is simply data that is interesting but it does not tell you anything about the future."

Participant from seismologist focus group 2: "With this picture, I do not know what information I need to understand. There are 23 earthquakes in January, but I don't know what I can do with this information."

Participant in public focus group 2: "For me it is interesting, if I was looking at the earthquake forecasting I would not go and see it but if it was something like an app, I live in Naples that is not that affected [by earthquakes], I would go and see it because it is interesting. I would go and see it not only for my location but also for others."

Providing some background knowledge on how the occurrence of earthquakes can in the short term increase the likelihood of future ones, might assuage some of these critiques and misperceptions. Still, though, it must be made absolutely clear that these capabilities do not allow for precise prediction of earthquakes in the future. It is possible that in Iceland, where the public understanding of geological processes is higher, this format would be considered more useful.

Based on the particular purpose of using recent past activity to understand why the forecast is higher, it might seem that recorded earthquake activity is most relevant for forecasts during seismically active periods rather than during periods of quiescence. Therefore one might argue that past activity should only be shown during an earthquake swarm, and that it could be removed from the quiescent time communication. However, one of the other purposes of the recorded activity, at least in the longer time frames shown during periods of quiescence, is to show that even though a forecast likelihood may be very low, earthquakes do happen in the location intermittently (or even, sometimes, in great numbers). Some participants did indeed seem to take this away from looking at the quiescent barchart and even described it as rather a revelation to see how much seismic activity there was in areas that they had considered earthquake-free.

Participant in public focus group 4: "There is rarely peace, when things seem to be quiet, something else arrives that makes you remember that one can never chill out."

- **Pattern-seeking in seismic activity**

There was an indication that some public participants thought a possible use of the historical data might be to search for patterns in the bar chart to allow them to make inferences about future earthquakes, and were frustrated when they couldn't clearly see such a pattern.

Participant in public focus group 2: "In 1980 there were 9 [earthquakes], then after 17 years there was another peak, and then we have to wait another 9 years to have another one. I do not find a pattern in this graph. It is informative and that's it. Functional only to satisfy your curiosity, not to prepare you."

Participant in public focus group 1: "The time ranging from 1997 and 2016 is 19 years, and 19 years before 1997 there are no earthquakes. I don't know why but I would expect another peak 19 years before."

To that extent, perhaps it was useful in helping correct misperceptions of cyclical patterns and predictability in earthquakes.

Other participants commented that the timeframe was too short to allow them to abstract anything into the future however (perhaps in part due to our making explicit that we calculated an "average week" for a location over a 40 year timeframe).

Participant in public focus group 3: "The average week is calculated based on 40 years, but here we have just a few weeks… When I look at this data, it is because there is high seismicity, but the time frame is too brief, so for me it is just to look at out of curiosity. What happened some weeks ago where I am? I look at it, but it doesn't give me either information on the future or help me with the average because that is calculated over 40 years."

Participant in public focus group 4: "I think that 200 or 300 years needs to be plotted to be able to quantify the likelihood there could be in the next years. From this little trend we are not able to do a projection."

One of the challenges with providing longer timeframes is that, due to space constraints, very long time frames need to be "binned" into larger categories, which changes the timeframe over which the data is displayed (e.g. from weeks to months or years, or years to decades or centuries) and thus the precision with which any trends (if they were present) might be identified. For the quiescent dashboard, this might not be such a problem since any credible trends would be expected to be macro-level anyway. Although identification of even macro-level patterns may mistakenly result in participants expecting to be able to make precise predictions from these trends (e.g. 'we seem to get earthquakes every 300 years, and we haven't had one for 300 years, therefore one will occur this year').

For a dashboard designed for use during a seismic swarm, if the communicator attempted to show a longer period of data and change the category bins, this may be more problematic. It is the fine-grained variations in day to day earthquake activity that can indicate the likelihood of another earthquake is higher. Again though, this also raises the issue of users expecting to be able to use the historical information to form precise predictions rather than mere increases in likelihood.

- **Misunderstandings about the difference between forecasts and recorded events**

When asked whether they would like to see historical forecast data, public participants in focus group 1 stated that they would, and that such information might help them check whether the forecasts are reliable and trustworthy.

Participant in public focus group 1: "You are giving me a recap on how to protect myself from earthquakes, so now I want to see if the forecasts I am looking at here could be reliable"

This suggests confusion between forecasts and actual recorded events, and also a lack of deep thought about the probabilistic nature of the forecasts. In order to check the reliability of probability estimates requires something similar to Brier scoring in weather forecasting (comparing actual number of events with forecast probabilities over suitable periods of time)- but which is very hard to do with low probability, rare events (but not impossible - methods have been devised).

In this case, the momentary confusion is understandable, and participants cannot be expected to have thought deeply about the formats and the data during a focus group where they are being shown new material on an unfamiliar topic.

6) Actions to take

In some of the early (alpha-phase) interviews, participants had expressed a desire for information to make the forecast actionable. Additionally, we were aware of the issue that was commonly raised when discussing OEF - that the absolute likelihoods of seismic events in forecasts were always too low or uncertain to use as a basis for any decisions about actions. Whilst this may well be true for high-cost actions such as evacuation, it is not true for low-cost actions such as testing emergency procedures, or making preparations. We hoped that highlighting some low-cost preparatory actions that could be undertaken might help users see the utility of the forecast and be deemed useful. See Figure 13 for the first iteration of this. We were aware that such information cannot always be provided by the same institutions that produce a seismic forecast, for legal reasons.

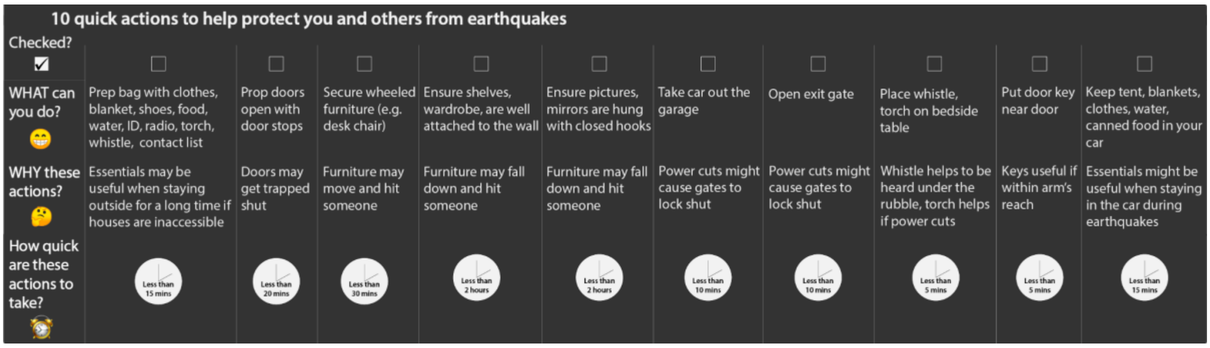


Figure 13: The 'actions to take' component of the dashboard

Some participants in the public focus groups felt this to be the most essential part of the dashboard, particularly where forecast likelihoods are higher. It was suggested that the actions could be made more prominent during such periods, or that their order could be changed to feature those more relevant to periods of higher risk. Other participants in both the public and seismologist focus groups suggested it would also be useful to have information on what to do during and after an earthquake too, or at least a prompt to review such information and/or a link to follow to view it.

- **Confusion about the purpose of the dashboard and this information**

There were indications across the public discussions of the actions that again, the purpose of the dashboard overall was not clear, with some again thinking it was a tool for alert or prediction rather than an informative tool. This could have been enhanced by the words "actions to help protect" in the title of this component of the dashboard, and the fact that preparation or preparedness were not mentioned. Indeed, certain participants discussed the actions as if they were to be taken during an acute period of risk and not as general preparatory actions, noting how if they perform one of the actions, they might not have time for the others (i.e. as if they were time pressured by an emergency event). In a similar vein, the estimated time taken for each action was criticised on the basis that it does not take into account how the emotion during an earthquake event might affect ability to act, again indicating a misunderstanding that the dashboard is a tool for use in an emergency.

Participant in public focus group 4: "The timing given to each action is poor because when a person experiences an earthquake they are very emotional and confused, they cannot remember where they put their shoes, at that moment their brain is so much in panic that it is not able to remember anything."

- **Misunderstandings about the probabilistic nature of the forecast**

Other public discussions regarding the actions related to the other, familiar, misunderstanding regarding the purpose of the dashboard - that it provided a prediction of earthquake occurrence rather than a probabilistic forecast. More specifically, some participants appeared to misunderstand the concept of "within the next week" meaning the forecast showed the likelihood of an earthquake occurring at any point within the next 7 days, instead believing it was forecasting an earthquake to occur specifically at the end of the next week, indicated by their thinking that they had a week in which to perform all the necessary actions to take.

Participant in public focus group 4: "In theory you have a week's time to do these things because it is the forecast in a week's time, isn't it?"

- **Comments on the visual presentation**

Regarding specific characteristics of the "Actions to take" design, some public participants found the use of emojis inappropriate for such a serious issue. Others commented that the reasons for the actions to take, whilst useful, could be included as a further layer or dropdown explanation for interested users, reducing visual clutter on the actions component. Yet others suggested pictograms might be a useful addition to the actions list. Several of the participants in both the public and seismologist focus groups did not like the visual feature of the checklist, being unsure what the empty boxes were actually for, and criticising how they couldn't actually be checked. Some also doubted whether they would actually check the boxes even if that feature were available. One public participant thought the boxes made it seem like a survey not a list of actions to take. While some public participants liked the timings given for each action, others from the both the seismologist and public focus groups felt that it was not possible to put a precise time on each action, as it would vary depending on other characteristics such as the size of someone's house. Some seismologist participants felt that whilst the short time estimations for some actions might encourage people to perform them, giving the time taken for other, lengthier actions, might discourage people.

- **Legalities of implementation**

A final but important point to make about the Actions to Take section of the dashboard regards the legalities of implementation. Not only would it require a systematically curated and officially approved list of possible preparedness actions, but approval would also be given their communication. Several countries have limitations on who can communicate such information, for example in Italy only Italian Civil Protection have this approval. Information relating to protective actions that could be taken can be linked to from any web page, so in future iterations of our design that is what we test, however we hope some of the above insights will be useful to those with the mandate to communicate this information.

Next iteration of the design

After synthesising the feedback on the dashboard, we created a new design incorporating as many improvements as we could to overcome some of the most common difficulties we encountered in the previous round of focus groups.

One of the big, general issues that we decided not to address explicitly in the new designs was trying to make the purpose of OEFs. We could have introduced text to give further explanations, and drafted some, but decided to sacrifice this in favour of greater simplicity and clarity in the interface. Once people become familiar with a concept and a format, through repetition, then explanation is unnecessary and would clutter a page, so we anticipated a separate page of text as an introduction. However, we wanted to test the new format without giving any extra introduction than participants in the previous round had, to see whether the formats of the information themselves could help with understanding the conceptual basis of the information.

The general issue that we did put a lot of thought into was how to help people understand instinctively the probabilistic nature of the forecasts. We considered where people come across probabilities (particularly quite low probabilities) in daily life and understand them: this led to us considering sporting odds, and the spin of a roulette wheel. We used these concepts to test trying to communicate low probabilities as long odds, and a graphic based on the design of a wheel.

Other key overall points:

- Using interactivity to reduce information overload, allowing formats to be displayed one at a time
- Minimising numbers to focus attention on the key figure (including giving a forecast for only one set magnitude of earthquake)
- Changing to communicate magnitude rather than intensity, using a different colour scale for it, and attempting to better communicate, graphically and verbally, the concept that the forecast is for earthquakes of ‘magnitude 4 and above’.
- Including some familiar example earthquakes on the magnitude scale to provide reference, to help people interpret what the different magnitudes mean in terms of subjective impact (rather than descriptions of the scientific definition of the magnitude)

Front page

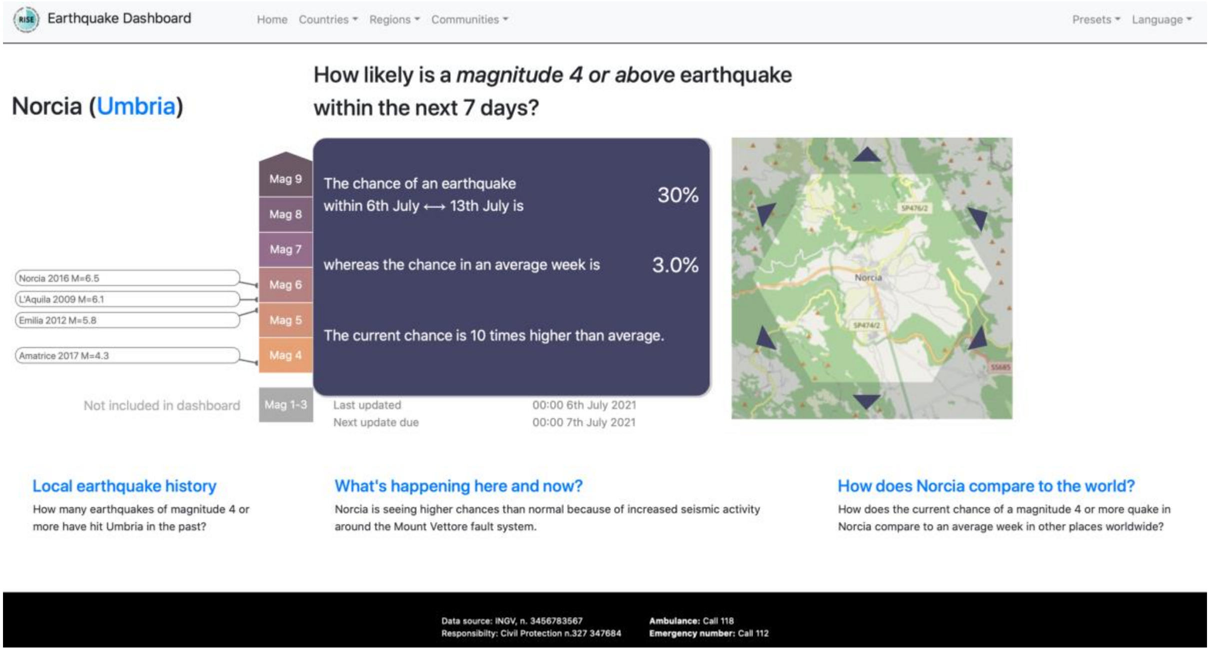


Figure 14: The front page of the second iteration of the dashboard (now interactive), incorporating a map alongside information from what had been in the previous Table and Text versions

Key points:

- Testing a few different ways of putting the absolute risk into meaningful context (one relative risk - compared with an 'average' week, but also trying options using the familiar format of betting odds which might help people more instinctively recognise the probabilistic nature of the forecast as many people are familiar with 'odds' in sport representing the chances of something happening or not)
- Using a single sentence to try to help the audience interpret the risk ('The current chance is x time higher than average'), which we also hope will remove concerns about what an 'average week' means.
- Trying alternative graphical representations, designed to help people interpret the probabilistic information (not shown in the example above)
- Providing links to the additional contextual information (temporal, using bar chart; risk comparisons, using risk ladder) clearly labelled to signpost to audience the intention of the information
- Removing the seismic information on the map in order to use it solely for orientation to the area specified in the forecast, and showing the grid for which the probabilities have been calculated. The map is also interactive and allows people to move from one grid square to another.
- Removing the use of a colour scale (previously saturations of blue) to indicate different likelihoods, to avoid potential confusion with the colours used for magnitudes.

Risk ladder ('How does X compare to the world?')

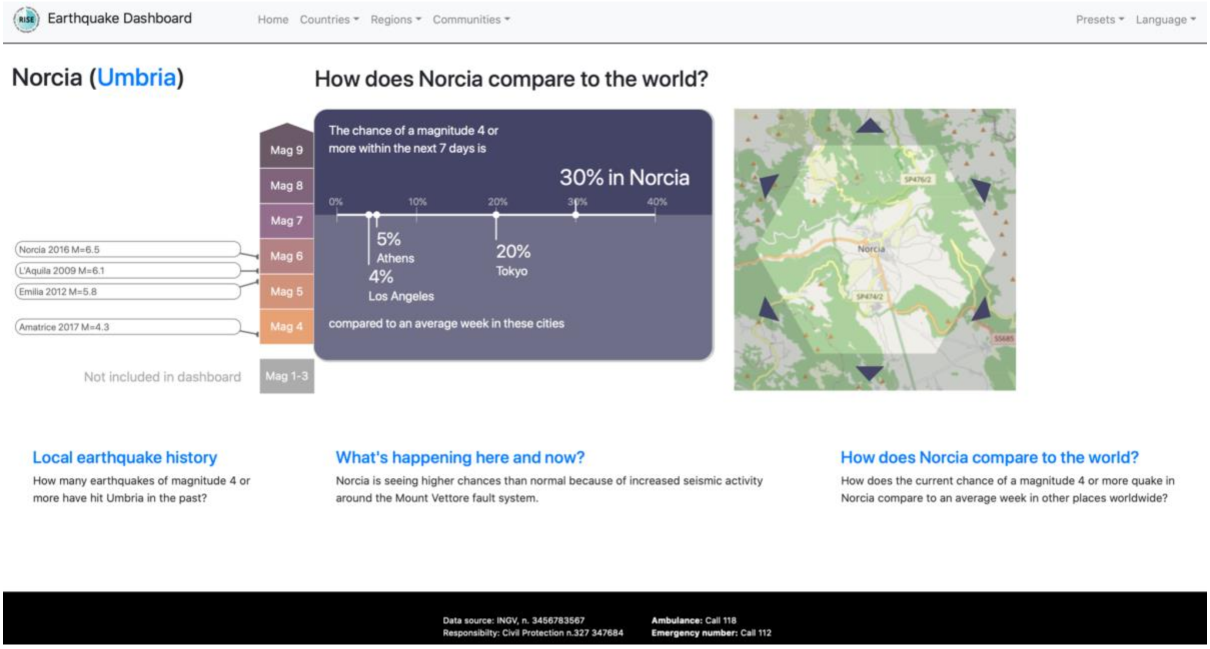


Figure 15: Second iteration of the Risk Ladder component of the dashboard, showing a hypothetical high probability forecast as if during an active earthquake sequence.

Key points:

- Changing the labelling and description to help guide the audience to the purpose of the comparisons
- Altering the design to make it clearer that the current forecast in the current selected location is 'on one side' and the comparator cities 'on the other'
- Removing colour scale on ladder

Bar chart ('Local earthquake history')

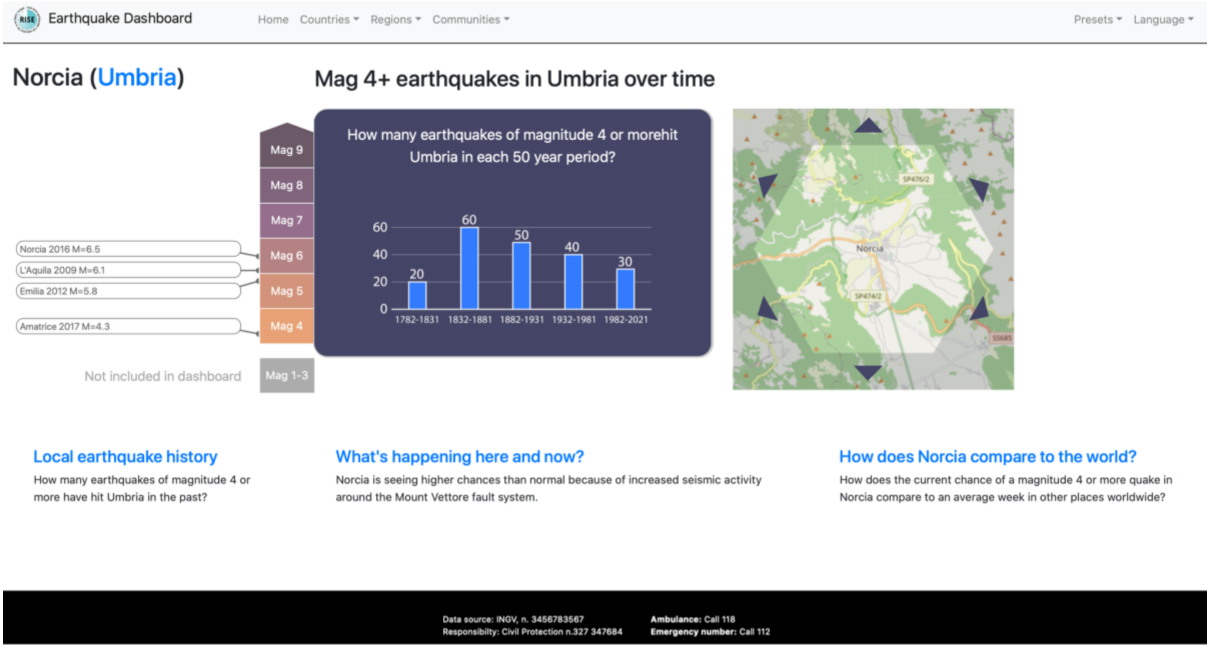


Figure 16: Second iteration of the 'Bar Chart' component of the dashboard, in a hypothetical quiescent period, where large periods of time are binned for each bar to represent.

- Changing the labelling and description to help guide the audience to the purpose and content of the historical data (e.g. making it clearer that the records are for the whole region)
- Using words in the titles and graphics (the magnitude scale) to attempt to make it clearer that the graph shows only magnitude 4+ earthquakes

Next steps

The new, interactive, dashboard is currently being tested with a variety of different audiences through interviews in Italy, Switzerland and Iceland. In each country, the participants will see a locally-adapted (i.e. showing local cities and with familiar national earthquake examples to illustrate different magnitudes) and translated version.

Large-scale quantitative testing, measuring the effect of different design features and formats on risk perception will be launched imminently to inform the final recommendations.

Earthquake Early Warning

One strategy to increase society's ability to take protective actions during shaking is the implementation of earthquake early warning (EEW) systems. Three global initiatives effectively drive these developments, namely the Sendai Framework for Disaster Risk Reduction, the Paris Agreement and the Sustainable Development Goals. Currently, EEW systems are operating in nine countries and being tested for implementation in thirteen countries (Cremen & Galasso, 2020).

The primary aims of EEW systems are to notify the general public about imminent strong ground shaking so that they can protect themselves on the spot and to trigger automated shutdown or safety procedures such as slowing down trains and securing critical infrastructure (Allen & Melgar, 2019). In recent years, several international research groups have assessed how the public perceives EEWs and what actions are triggered by the warnings (Becker et al., 2020; McBride et al., 2020; Nakayachi, Becker, Potter, & Dixon, 2019; Sutton, Fischer, James, & Sheff, 2020; Tan et al., 2021). We contribute to this investigation by exploring the public's expectations and needs in European countries and also countries where damaging earthquakes are expected only every 50 to 150 years, e.g. Switzerland.

So far, we have conducted several expert interviews with seismologists, social scientists and practitioners working on EEW systems. In addition, we conducted semi-structured interviews with the public in Iceland, Switzerland and Italy to get a first impression of the public's attitudes towards EEWs. Based on these insights, we designed several EEW messages and refined them with EEW experts. Afterwards, we tested these designs with the Swiss public (n=596), using a between-subject experiment survey. The designs are listed and the results are summarized in the following sections.

Insights from the interviews with the experts

We conducted interviews with EEW experts working in the US, Japan, Mexico and Italy. In the latter, the EEW system is being tested and in the other three the systems are already operational. In Table 1, we summarize the main insights gained from the interviews.

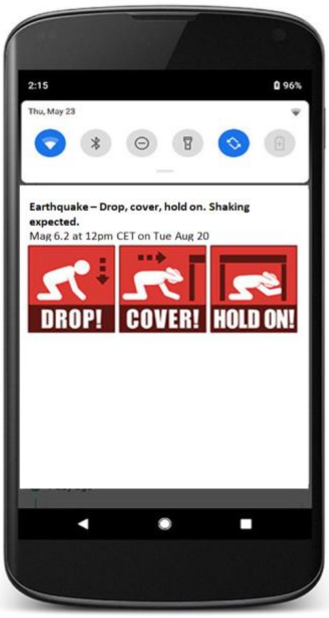
Table 1: Main insights from the expert interviews regarding EEWs

Purpose of EEW	**Format & Medium of EEW**
● Alert people of an impending earthquake ● Mitigate earthquake-related damages ● Exchange data with the community	● **Format:** maps (shake maps; interactive maps), graphs, descriptive text, layered communication ● **Information:** As a first message**,** the affected area, the time and the expected shaking intensity are communicated. As a second message, the magnitude, the depth and the expected damages are communicated. ● **Medium:** Radio, TV, loud speakers, phone app (e.g. SkyAlert), social media and machine-to-machine

	processes.
Audience	**Future potential**
• General public • Specific audiences (engineers) for specific information • Operators of critical infrastructures (automatic activated safety procedures)	• Improving the distribution time of alerts by using cheaper and more detectors. • Improving the accuracy of seismic intensity estimation.
Challenges	**What has already been tested?**
• **Warning time:** Warnings may not arrive in time before the shaking starts [blind zone]. • **Aftershock sequence**: Many alerts are sent to the public within a short period of time. • **Limited accuracy:** The predicted intensity is not exact, and thus, the information on the warning and what users actually experience can differ (e.g., earthquake in LA in 2019). • **Retentions**: App uninstalled if no further earthquakes occur. • **Thresholds**: People have different preferences to receive alerts. Some of them want to receive EEWs for all earthquakes that are felt and others only for the ones that may cause damages. • **Information update:** The precise numbers (e.g. magnitude) may change within some minutes after the earthquake. • **Scales:** Different intensity scales may lead to confusion (e.g. tourists in Japan). • **False alerts**	• **Preferred media**: mobile phones (especially apps), TV, radio, loud speakers and social media. • **Preference for accurateness and speed** • **Reasons for interest:** worries about loved ones, discover the quake severity, out of curiosity • **Triggered behaviours**: People tend to mentally prepare instead of taking active actions such as taking cover under a table.

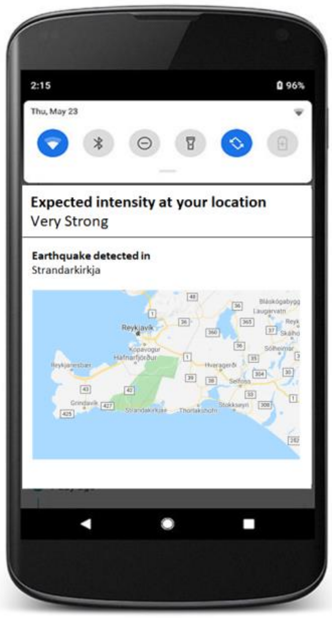
Insights from the interviews with the public

During the interviews with the public we presented three EEW message designs to the public and asked them to evaluate them (Figure 17Figure 17: The three EEW messages we showed to the interviewees in Iceland. The ones for Italy and Switzerland). The three messages differ regarding the information elements. The first one contained pictograms, the second one had a countdown and the intensity level as prominent information and the third one a map that depicted the epicenter. In Table 2 we summarize the public's preferences and needs and some implications for the design of the messages.



Pictograms in the focus.

Intensity and arrival time in the focus.



Map and thus affected area in the focus.

Figure 17: The three EEW messages we showed to the interviewees in Iceland. The ones for Italy and Switzerland

Table 2: Public's preferences and needs and design implications with regard to EEWs.

Public's preferences and needs	**Design implications**
• Behavioural recommendations (take cover under a table) as pictograms • Arrival time as a count down if it is feasible • Information about epicenter and distance from one's location • (Trusted) information source • Link to further information • Intensity map • Verbal expression of the intensity/magnitude level • Magnitude • Two messages: First one should be short and simple and the second one with more detailed information. • Message accompanied by a sound.	• Use "red" to stress that it is urgent • Clear and simple title, e.g. "Attention Earthquake - Strong Shaking Expected". • Text not too small • No abbreviations

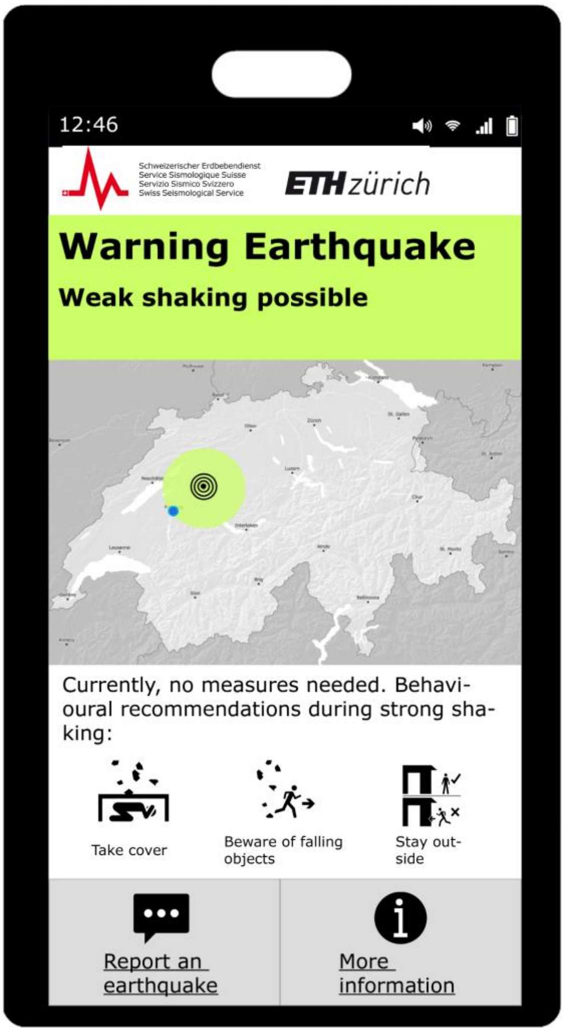
As mentioned above, we conducted the interviews in Switzerland, Italy and Iceland. We thus also compared the preferences between these countries. First, Icelanders said that they know what to do during the shaking and thus do not need the behavioural recommendations on the message. In comparison, people in Switzerland and Italy want pictograms depicting what one should do during strong shaking. Second, people in all three countries struggle with the abbreviations (e.g. Mag, CET, Roman numbers). Third, people in Italy who have already experienced strong shaking think that they would panic if they received such a warning. This was neither mentioned by the Swiss nor the Icelandic people. Fourth, the majority of the people in all three countries think the map should not be part of the first EEW but rather be embedded in the second message after the shaking is over.

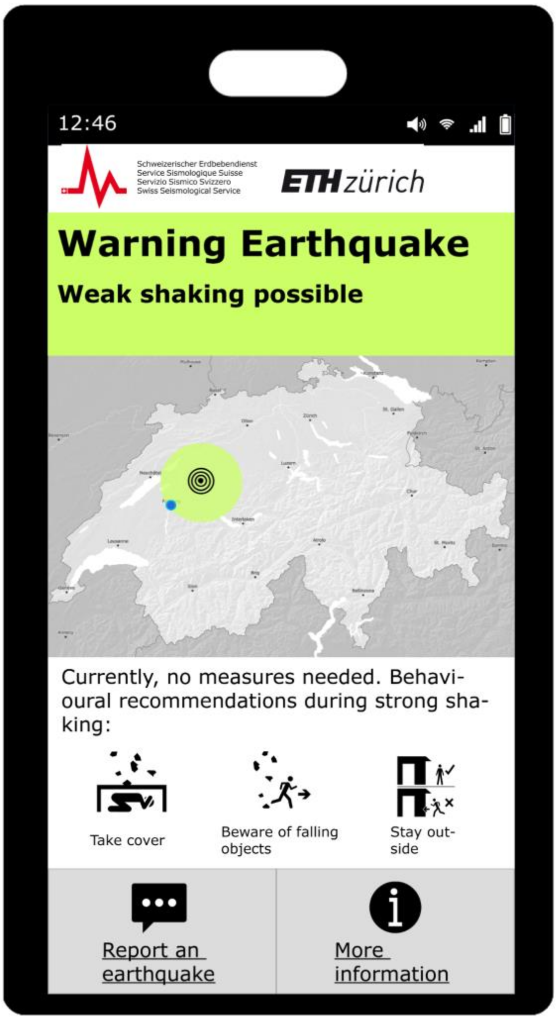
Designs

The EEW and REI we designed based on the insights from the expert and public interviews are presented in Table 3 and

Table 4. In order to guarantee scientifically high quality products, a group of experts provided their inputs and gave feedback on these EEW and REI messages. One relevant insight from the discussions was to not include considering the limits of EEW in Switzerland. In most cases, people in Switzerland will receive the EEW only after the shaking has already started or even when it is already over. We thus designed test messages that are feasible and realistic in these circumstances. In addition, previous studies in Switzerland showed that most people do not know what to do during an earthquake, in consequence the design focuses on clear and short instructions including pictograms. The messages were then tested with a survey (see next section).

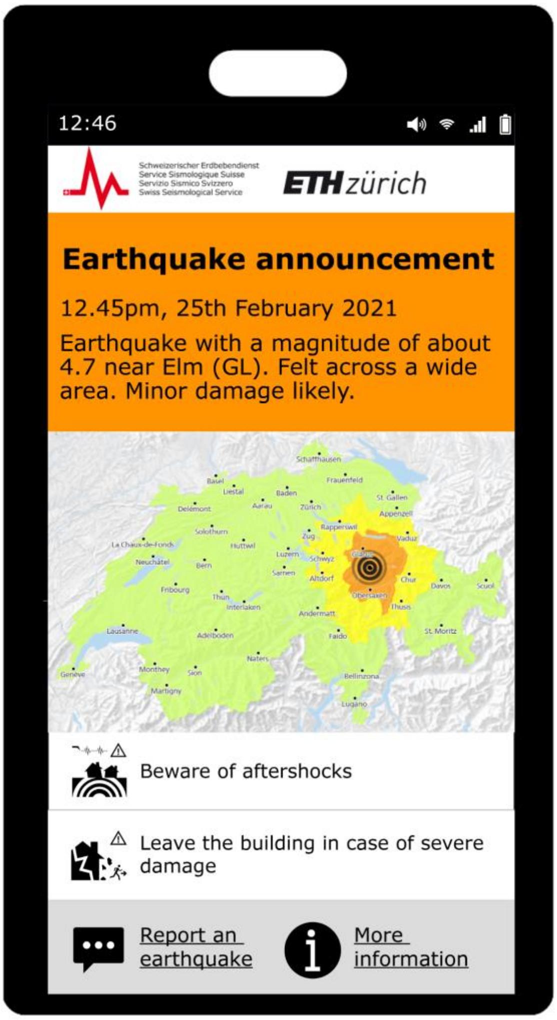
Table 3: EEW messages designs

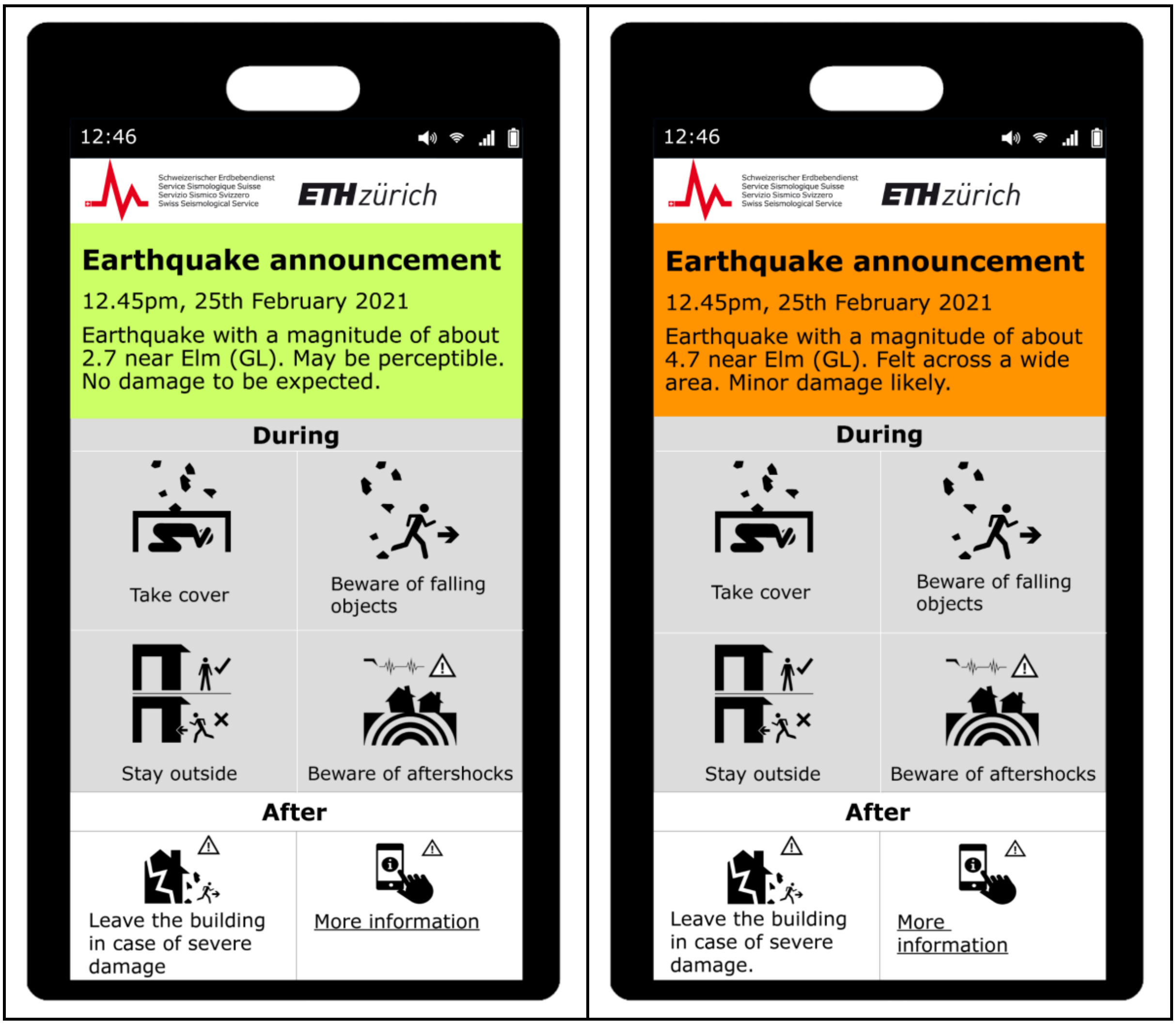
Weak shaking message with map	Strong shaking message with map
	
Weak shaking message with pictograms	Strong shaking message with pictograms



Audio weak shaking message	Audio strong shaking message
Attention earthquake! Weak shaking possible at your location! Be prepared.	***Attention earthquake! Take cover. Strong shaking possible at your location.***

Table 4: REI messages

<table>
<tr><td>Weak shaking message with map</td><td>Strong shaking message with map</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Weak shaking message with pictograms</td><td>Strong shaking message with pictograms</td></tr>
</table>



Insights from the survey

To test the different EEW designs listed in Table 3 and Table 4, we conducted an online survey with two between-subject experiments in Switzerland's German- and French-speaking parts in March/April 2021 (n=596). To this end, the participants were randomly assigned to one EEW message and one REI message. Regarding the EEW message, half of them saw the message only for 3 seconds and the others for 5 seconds. This allowed us to assess how much information can be read and remembered in such a short time. In addition, we put the participants in a "stress situation", mirroring such a situation in the best possible way.

In Table 5, we summarized the main results divided into the public's general preferences, the EEW message design, the purpose of the second message/REI and challenges.

Table 5: Results from survey with general public

Public's general preferences

- The Swiss public prefers to receive EEWs for earthquakes that are at least felt.
- The preferred warning time is 20 or more seconds, however technically feasible are about 3-5 seconds. In some areas of the world, longer warning times are possible.
- The Swiss public wants to receive the EEWs as push notifications on their mobile phones, supported by radio and public announcements (e.g. sirens).

Message Design

Overall, all the messages we designed were well perceived and rated as understandable, useful, trustworthy and informative. However, we have some implications about which elements trigger which behaviours:

- Pictograms trigger people to take actions such as protecting oneself on the spot, and are perceived as understandable, useful, trustworthy and informative.
- Maps trigger people to warn and protect others, and make it clear whether they are personally affected or not.
- Audio messages trigger people to protect themselves on the spot or to mentally prepare. In addition, people recall most of the information they heard.
- The red and bold highlighted information at the top and middle of the messages is better recalled than the grey, smaller message at the end of the message.
- Strong shaking messages motivate people to take actions compared to weak shaking messages. Thus, they react proportionally to the intensity.

Purpose of second message

- The REI messages for strong shaking trigger people to move nearby to where they think they are safe, to look for further information or to share the information with friends/family. In comparison, weak shaking messages do not trigger any actions.
- Considering people's cognitive capacities, not too much information should be included in the message. People ask for general information about the earthquake, behavioural recommendations during the shaking and first information about possible aftershocks. Information that could be accessed via a link are behavioural recommendations for after the shaking, information about the extent of the damages/impacts, secondary hazards and more details about the expected aftershocks.
- When the messages are sent via an app, interactive features such as sharing functions, report buttons or chat forums can be an added value. These features support people to handle the (emergency) situation.

Challenges

- Experience bias: Some people think that they do not need an EEW because nothing happened last time so nothing will happen during the next earthquake.
- Lack of knowledge: Some people do not know what actions to take during an earthquake.
- Limited warning time: A majority thinks that they will not have enough time to react.
- Misconception: Many people think they know how an EEW system works but actually they do not.

Information campaigns are needed to address these challenges and to clarify the potentials and limits of an EEW system.

Next steps

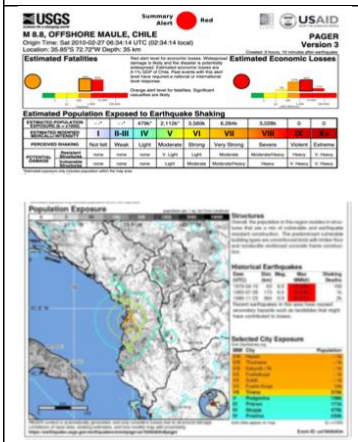
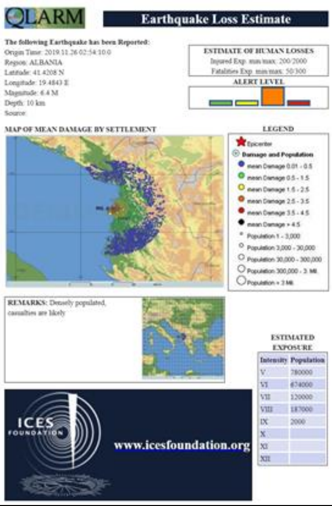
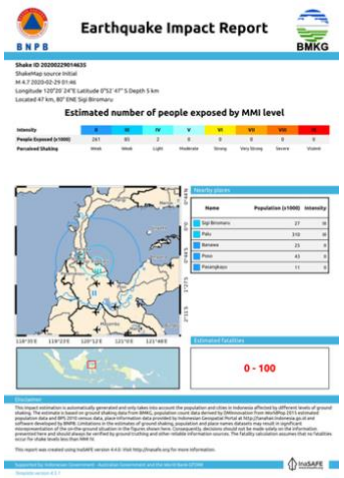
The results of the survey will be peer-reviewed, published, and presented at conferences. In addition, we aim at conducting a similar survey in Italy to provide further cross-cultural comparisons. With the time limit of the message display, a quasi-realistic setting has been established for this study. Nevertheless, people might react differently when expecting or experiencing an earthquake. Therefore, future research should evaluate the message designs implemented in operating EEW systems.

Rapid Impact Assessment

Rapid Impact Assessment (RIA) after a severe earthquake can support civil protection agencies and emergency services to rapidly gain an overview of the expected building damages, number of fatalities, injured and displaced persons as well as economic losses. Such information allows coordinating and allocating the resources for the emergency response in an efficient manner. Of course, similar outputs can also be produced in advance, whereby these scenarios can be used to build up and support the awareness for damaging earthquakes among different stakeholders and help emergency managers to define disaster management plans. We use our close involvement in the development of various products to communicate the results of the European Seismic Risk Model 2020 (ESRM20) and the Swiss Seismic Risk Model to explore and test different aspects.

Insights from the interviews with the experts

As a preliminary study, we conducted interviews with experts working on RIA. We had a look at the following products: PAGER (national and global), ShakeCast (national), QLARM (global), InaSAFE (national) and Globale Dynamic Exposure Model (global, still in development), see . The insights of these interviews fed into the design of the output products of the ESRM2020 and the Swiss Seismic Risk Model. The results of the interviews are summarized in Table 6.

QLARM	USGS PAGER	InaSAFE
		

The generation of the rapid impact assessment report consists of several steps. When an earthquake occurs the sensor network registers the seismic waves and the shakemap is generated. The data of the hypocentre and magnitude is then fed into the risk-assessment models together with the population and building data. Either global or national data sets with different resolutions are used. The model then generates the report that is published on the website and sent to specific stakeholders. Moreover, some agencies also take the feeds for the report and calculate on their own further potential damages. When the data of the shakemap changes the report is newly generated. However, only when the alert level changes the subscribers again receive an alert in the US. Or in Indonesia, for major earthquakes a second report is generated.

Up to now, many rapid impact assessments are publicly available and therewith open to use for everyone. It is well known that RIA's are frequently picked up by media. However, from the interviews we learnt that the providers of RIA primarily target professionals users such as first responders (e.g., civil protection) or international emergency agencies. In the US and Indonesia, for example, the report is directly sent to the disaster agency operations room/ national watch office of FEMA. However, experts agree that also the public and media is accessing e.g. PAGER assessments and emphasize that they should have access to general overviews. In addition, stakeholders with specific requirements (e.g. insurances, critical infrastructure operators) should or already have (e.g. US) access to more detailed information and data which they can include into their models. In any case, in particular in countries where damaging earthquakes only occur rarely, even when restricting the access to RIA products, the primary users will not be very familiar with the outputs. Therefore, RIA outputs need to be assessed and designed to be as understandable and accessible as possible. We argue, that already today RIA's are not only reaching well-trained professionals and therefore should be tested with a broader audience to avoid misunderstandings and misconceptions.

Table 6: Insights from the expert interviews

Purpose of RIA's	**Format**
• Rapid loss assessments after an event to have a first overview of the damages, fatalities and people in help. • Scenarios of hypothetical or historical earthquakes to define disaster management plans. Quote: *"We don't claim to estimate, to predict losses at a facility level but we can allow people to prioritize where within this very small region the worst situations are gone be."*	Content • Magnitude, location, depth, time • Intensity or mean damage map • Estimate of fatalities and economic losses • Histogram with alert levels • Table with affected cities Communication means • Website, E-mail, Feeds, SMS, Social Media Quote: *"[...] I spent a lot of my time talking to these folks, which is really key because we are not only addressing their concerns but we then fall these concerns back on how we deliver things."*
Audience	**Future Potential**
• First responders • (Inter)national emergency agencies • Insurance companies • Critical infrastructure owners • General public • Media reporter	• Going from global to local (higher resolution data) • Include data about hospitals & schools • Also inform about hazards triggered by the earthquake • More detailed national risk assessment reports • Combine it with statements about aftershock forecasts

Challenges	What has already been tested?
● No second reports for major aftershocks ● Updated reports can lead to confusion ● Communication of the prediction, likelihood, uncertainties ● Privacy issues/ Disclaimer ● Lack of time and resources for the maintenance of the products and 24/7 contact possibilities. ● Normalization bias ● Media interpreting the numbers not correctly.	● Scientists' experience & expertise ● Informal feedback from the primary users ● Workshops ● Scientific meetings ● Indicate the number of fatalities as ranges so that the media does not report an exact number that may be with high probability not correct.

What is still needed is how one can best communicate RIA including the uncertainties behind the models and final numbers in the report. One participant stated: *"But you never have to show the exact medium because that number is always gonna be wrong. And if you are wrong, that is what people focus on. But if you have a histogram you are always right, it's always in there somewhere. And the histogram allows people to understand how uncertain the estimate is. So, that seemed to work out pretty well. But, it's still not clear what fraction of the population fully understand that but they do get the alert level, it's green, yellow, orange or red."* We thus will test different RIA formats and explore different ways in portraying uncertainties with different user groups.

The European Seismic Risk Model

In 2021, a fully open access risk model for Europe will become available. Different products and outreach materials are drafted to inform various stakeholders and the interested public about the model's results. This offers a unique opportunity to test communication materials focusing on seismic risk. For the different target audiences, we design(ed) various products in order to meet their needs: website, web-viewer, detailed report, factsheet, FAQ, flyer, web material, standard presentation, poster and an infographic. In addition, events such as webinars, workshops and information events were/are organized to inform the different target audiences about the new European risk model.

In the first step, we have conducted an interactive online survey testing the web-viewer of the risk model with potential end-users (see next section).

In a second step, we will test different map designs and the understanding of key concepts and core messages using first drafts of a risk poster, showing the most important elements of the model. The poster design will be tested through a survey with students from different European countries to ensure that the relevant information is understood and the communication goals are met.

Recommendations for the design of such products: i) ensure an appropriate balance between text and visual elements (not too much text); ii) use a corporate design to make the different products consistent; iii) use colour-blind-friendly palettes; iv) define the key messages thus the information that should be communicated before starting designing the products; v) use icons and pictograms to graphically present scientific concepts; and vi) let experts check the text on the poster.

Web-viewer of the risk model

To assess the user needs and explore the usability of the risk web-viewer, we conducted an interactive online survey with a total of 17 participants. The participants were researchers, civil engineers, cat risk modellers or persons from the civil protection agency. The web-viewer can be accessed via the following link: https://maps.eu-risk.eucentre.it/map/europe-risk-level-0_beta-version/#5/41.856/7.752 (see Figure 18). In Table 7, we summarized the practical implications we gained.

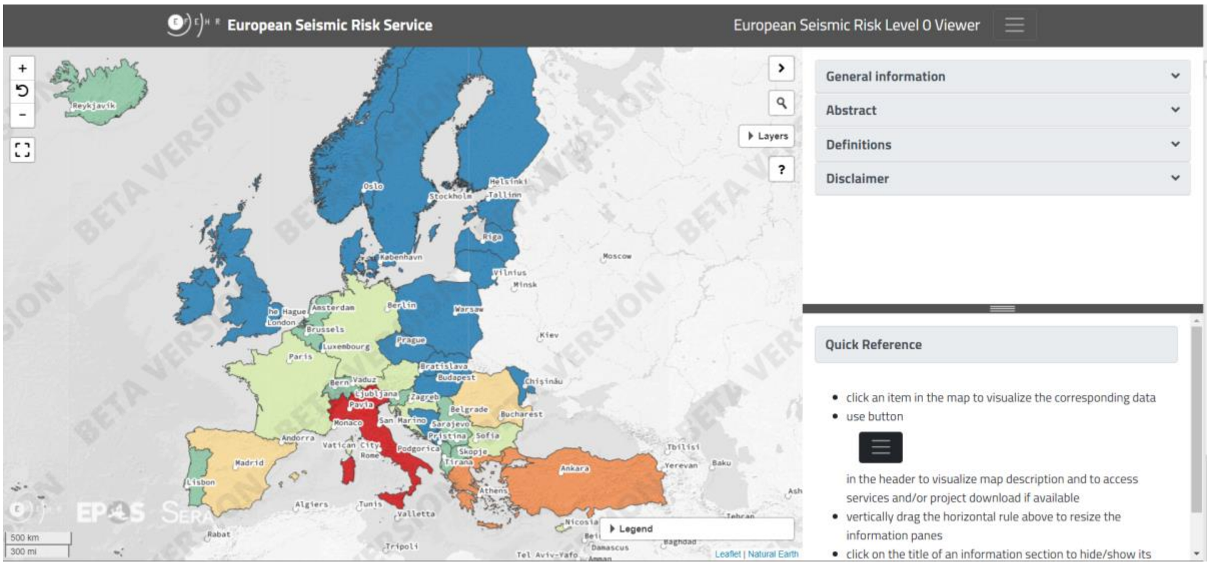


Figure 18: Interactive web-viewer of the European Seismic Risk Model 2020 [screenshot], accessible here: https://maps.eu-risk.eucentre.it/map/europe-risk-level-0_beta-version/#5/41.856/7.752

Table 7: Summary of the practical implications for designing an interactive web-viewer for risk models.

Preferred information
• Direct and indirect economic losses • Number of casualties, fatalities and people in help • Damages on physical assets (e.g. damages, infrastructure) • Hazard and exposure (population + building) data • Fragility and vulnerability models for residential and commercial buildings • Social vulnerability or resilience indicators • Building stock information/ Differentiation between commercial, industrial and residential buildings • Uncertainties associated with the models
Map preferences
Preferred maps • Map of average annual loss (M EUR) > Map of average annual loss ratio (per mille) > Map of the 200 years return period loss (M EUR) • Direct access to hazard and exposure map • Mapping of social vulnerability indicators • Map of the distribution of the collapse risk of buildings

- Ability to download maps as .csv
- The risk results covering both economic losses and fatalities should be provided together in the same interactive map viewer.

Preferred resolution

- Gridded map [e.g. 1km x 1km] > National level > NUTS19
- Resolution: countries, cities and municipalities
- The risk results for all levels of resolution should be included in the same map viewer so that all information is together.

Preferred layers

- Populated places / density of the population
- Significant earthquakes [according to the NCEI WDS database]
  → By clicking on the event, more detailed information about the earthquake pops up.
- Active/major faults
- Return periods: 50, 200 and 500 years
- Relevant infrastructure/lifelines
- Shaded relief not wished

Purpose of the risk model

- To give estimates of risk levels at various return periods for the mapped economic exposure.
- To provide an overall view of seismic risk in Europe and to compare seismic risk in the different EU countries.
- To guide the development of public/private risk mitigation strategies of all sorts, such as deployment of wide-scale structural upgrading campaigns.
- To compare with and improve existing vendor models of European seismic risk.
- To provide easy access to specific risk metrics for the whole Europe accompanied by the data/models used for its development.
- To raise awareness within the scientific and engineering communities.
- To provide reliable data that can be quickly found.
- To homogenise the seismic hazard maps along the boundaries of the European countries.
- To increase awareness of seismic risk in Europe at the levels of both the government and the public.
- To estimate the number of displaced people and potential casualties, as part of the national disaster management plans (preparedness phase).

Design implications

- The web-viewer was overall rated as easy to navigate, attractive, clear, informative and useful.
- Only the topographic layer was not well visible and had to be adjusted.
- The use of clear and understandable legends are important, i.e. provide the same information in the legend and the information box .
- Further information (pop-up windows) have to be intuitively found, i.e. using clear icons.
- When one has an information box on the side of the map, make sure that it is obvious how to open and close it.

The Swiss Seismic Risk Model

For 2022, the release of the first national seismic risk model for Switzerland is planned. An internal group at the Swiss Seismological Service at ETH Zurich has, in parallel to the model development, started to work on communication products. Drafts for RIA's and scenarios have been developed first and will now be tested with the main target audiences - cantonal authorities, first responders and civil protection - to check whether the information is presented in an understandable and clear way and whether relevant information is currently missing. To this end, we conduct(ed) workshops with the target audiences, where we present(ed) the main models and products and gather(ed) their feedback. Based on the feedback, we will then adjust the products accordingly. We here summarized the results in form of recommendations of the workshops we conducted so far:

Recommendations for the design

- Ensure that the information display is consistent with already known and established products.
- Add a disclaimer that the information may be updated and/or changed at any time.
- Indicate the source of the information to establish trust.
- Ensure that the alert levels are consistent with the information on other national platforms. In Switzerland for example all natural hazards are divided into five categories. For earthquakes, the classification depends on the intensity of the earthquake. This classification was applied for the risk product too.
- Choose reasonable thresholds for the categorization of the number of fatalities, displaced, injured, damaged buildings and economic losses.
- Provide information not only on a national but also on a cantonal or even communal level.
- Verbally describe the intensity levels which are usually indicated as Roman numbers.
- Provide short explanation texts about how to read the visualizations.

General recommendations

- Different product views for the public/media and the cantonal/national authorities
- Information needed: map with the impacts, number of fatalities and injured, number of damaged buildings and economic losses
- Additional information: local intensity, expected aftershocks, vulnerability of the most relevant infrastructure in a region, secondary hazards
- Support from experts during the emergency needed
- Risk comparison between different regions is appreciated
- Tool which allows to calculate the risk for a specific building type is wished.
- Have different thresholds for the specific regions (e.g., in Switzerland cantonal thresholds)
- Use already established communication means to distribute the rapid impact assessment report.

References

Allen, R. M., & Melgar, D. (2019). Earthquake Early Warning: Advances, Scientific Challenges, and Societal Needs. *Annual Review of Earth and Planetary Sciences*, *47*(1), 361–388. https://doi.org/10.1146/annurev-earth-053018-060457

Becker, J. S., Potter, S. H., McBride, S. K., H. Doyle, E. E., Gerstenberger, M. C., & Christophersen, A. (2020). Forecasting for a Fractured Land: A Case Study of the Communication and Use of Aftershock Forecasts from the 2016 Mw 7.8 Kaikōura Earthquake in Aotearoa New Zealand. *Seismological Research Letters*, (November). https://doi.org/10.1785/0220190354

Becker, J. S., Potter, S. H., McBride, S. K., Wein, A., Doyle, E. E. H., & Paton, D. (2019). When the earth doesn't stop shaking: How experiences over time influenced information needs, communication, and interpretation of aftershock information during the Canterbury Earthquake Sequence, New Zealand. *International Journal of Disaster Risk Reduction*, *34*(November 2018), 397–411. https://doi.org/10.1016/j.ijdrr.2018.12.009

Becker, J. S., Wein, A., Potter, S., Emma Doyle, & Ratliff, J. L. (2015). Aftershock communication during the Canterbury Earthquakes, New Zealand: Implications for response and recovery in the built environment. Retrieved from http://pubs.er.usgs.gov/publication/70148058

Brath, R., & Peters, M. (2004). Dashboard design: Why design is important. *DM Review*, *October*.

Brooks, B. A., Protti, M., Ericksen, T., Bunn, J., Vega, F., Cochran, E. S., … Glennie, C. L. (2021). Robust Earthquake Early Warning at a Fraction of the Cost: ASTUTI Costa Rica. *AGU Advances*, *2*(3), 1–17. https://doi.org/10.1029/2021av000407

Camerer, C. F., & Kunreuther, H. (1989). Decision Processes for Low Probability Events: Policy Implications. *Journal of Policy Analysis and Management*, *8*(4), 565–592. https://doi.org/10.3109/15569527.2012.667030

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370. https://doi.org/10.1007/BF00122575

Cohen, D. J., Ferrell, J. M., & Johnson, N. (2002). What very small numbers mean. *Journal of Experimental Psychology: General*, *131*(3), 424–442. https://doi.org/10.1037/0096-3445.131.3.424

Cremen, G., & Galasso, C. (2020). Earthquake early warning: Recent advances and perspectives. *Earth-Science Reviews*, *205*(March), 103184. https://doi.org/10.1016/j.earscirev.2020.103184

Dallo, I., & Marti, M. (2021). Why should I use a multi-hazard app? Assessing the public's information needs and app feature preferences in a participatory process. *International Journal of Disaster Risk Reduction*, *57*(March), 102197. https://doi.org/10.1016/j.ijdrr.2021.102197

Dallo, I., Stauffacher, M., & Marti, M. (2020). What defines the success of maps and additional information on a multi-hazard platform? *International Journal of Disaster Risk Reduction*, *49*, 101761. https://doi.org/10.1016/j.ijdrr.2020.101761

Denes-Raj, V., Epstein, S., & Cole, J. (1995). The Generality of the Ratio-Bias Phenomenon. *Personality and Social Psychology Bulletin*, *21*(10), 1083–1092.

Doyle, E. E. H., McClure, J., Potter, S. H., Lindell, M. K., Becker, J. S., Fraser, S. A., & Johnston, D. M. (2020). Interpretations of aftershock advice and probabilities after the 2013 Cook Strait earthquakes, Aotearoa New Zealand. *International Journal of Disaster Risk Reduction*, *49*, 101653. https://doi.org/10.1016/j.ijdrr.2020.101653

Few, S. (2006). *Information Dashboard Design: The effective visual communication of data*. (C. Wheeler, Ed.), *Information dashboard design: displaying data for at-a-glance monitoring* (First Edit). O'Reilly. Retrieved from https://books.google.de/books/about/Information_Dashboard_Design.html?id=7k0EnAEACAAJ&redir_esc=y

Field, E. H., Jordan, T. H., Jones, L. M., Michael, A. J., Blanpied, M. L., Abrahamson, N., … Wein, A. (2016). The potential uses of operational earthquake forecasting. *Seismological Research Letters*, *87*(2A), 313–322. https://doi.org/10.1785/0220150174

Freeman, A. L. J., Kerr, J., Recchia, G., Schneider, C. R., Lawrence, A. C. E., Finikarides, L., … Spiegelhalter, D. (2021). *Communicating personalized risks from COVID-19: Guidelines from an empirical study*. *Royal Society Open Science* (Vol. 8). https://doi.org/10.1098/rsos.201721

Gaspar-Escribano, J. M., & Iturrioz, T. (2011). Communicating earthquake risk: Mapped parameters and cartographic representation. *Natural Hazards and Earth System Science*, *11*(2), 359–366. https://doi.org/10.5194/nhess-11-359-2011

Gerstenberger, M. C., Wiemer, S., Jones, L. M., & Reasenberg, P. A. (2005). Real-time forecasts of tomorrow's earthquakes in California. *Nature*, *435*(7040), 328–331. https://doi.org/10.1038/nature03622

Gerstenberger, M., McVerry, G., Rhoades, D., & Stirling, M. (2014). Seismic hazard modeling for the recovery of christchurch. *Earthquake Spectra*, *30*(1), 17–29. https://doi.org/10.1193/021913EQS037M

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health Statistics: Toward an Evidence-Based Society. *Psychological Science in the Public Interest*, *8*(2), 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Goltz, J. D. (2015). A further note on operational earthquake forecasting: An emergency management perspective. *Seismological Research Letters*, *86*(5), 1231–1233. https://doi.org/10.1785/0220150080

Granger Morgan, M., Fischoff, B., Bostrom, A., & Atman, C. J. (2002). *Risk Communication: A Mental Models Approach*. Cambridge University Press.

Hagemeier-Klose, M., & Wagner, K. (2009). Evaluation of flood hazard maps in print and web mapping services as information tools in flood risk communication. *Natural Hazards and Earth System Science*, *9*(2), 563–574. https://doi.org/10.5194/nhess-9-563-2009

Halpern, D. F., Blackman, S., & Salzman, B. (1989). Using statistical risk information to assess oral contraceptive safety. *Applied Cognitive Psychology*, *3*(3), 251–260. https://doi.org/10.1002/acp.2350030305

Hayes, G. (n.d.). EQ Magnitude, Energy Release, and Shaking Intensity. Retrieved from https://www.usgs.gov/media/images/eq-magnitude-energy-release-and-shaking-intensity-5

Heckler, A. F., Mikula, B., & Rosenblatt, R. (2013). Student accuracy in reading logarithmic plots: The problem and how to fix it. *Proceedings - Frontiers in Education Conference, FIE*, 1066–1071. https://doi.org/10.1109/FIE.2013.6684990

Jordan, T. H., Marzocchi, W., Michael, A. J., & Gerstenberger, M. C. (2014). Operational earthquake forecasting can enhance earthquake preparedness. *Seismological Research Letters*, *85*(5), 955–959. https://doi.org/10.1785/0220140143

Jordan, Thomas H., Chen, Y. T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., … Zschau, J. (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Annals of Geophysics*, *54*(4), 319–391. https://doi.org/10.4401/ag-5350

Joslyn, S., & Savelli, S. (2010). Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteorological Applications*, *17*(2), 180–195. https://doi.org/10.1002/met.190

Kaplan, R. M., Hammel, B., & Schimmel, L. E. (1985). Patient information processing and the decision to accept treatment. *Journal of Social Behavior and Personality*, *1*(1), 113–120.

Keller, C. (2011). Using a Familiar Risk Comparison Within a Risk Ladder to Improve Risk Understanding by Low Numerates: A Study of Visual Attention. *Risk Analysis*, *31*(7), 1043–1054. https://doi.org/10.1111/j.1539-6924.2010.01577.x

Keren, G., & Gerritsen, L. E. M. (1999). On the robustness and possible accounts of ambiguity aversion. *Acta Psychologica*, *103*, 149–172. https://doi.org/10.1016/S0001-6918(99)00034-7

Kunz, M., Grêt-Regamey, A., & Hurni, L. (2011). Customized Visualization of Natural Hazards Assessment Results and Associated Uncertainties through Interactive Functionality. *Cartography and Geographic Information Science*, *38*(2), 232–242. https://doi.org/10.1559/15230406382232

Lipkus, I.M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute. Monographs*, *27701*(25), 149–163. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024191

Lipkus, Isaac M. (2007). Numeric, verbal, and visual formats of conveying health risks: suggesting best practices and future recommendations. *Medical Decision Making*, *27*(5), 696–713. https://doi.org/10.1177/0272989X07307271

Marti, M., Stauffacher, M., & Wiemer, S. (2019). Difficulties in explaining complex issues with maps. Evaluating seismic hazard communication - the Swiss case. *Natural Hazards and Earth System Sciences Discussions*, (July), 1–32. https://doi.org/10.5194/nhess-2019-112

Marzocchi, W., Taroni, M., & Falcone, G. (2017). Earthquake forecasting during the complex Amatrice-Norcia seismic sequence. *Science Advances*, *3*(9), 1–8. https://doi.org/10.1126/sciadv.1701239

McBride, S. K., Bostrom, A., Sutton, J., de Groot, R. M., Baltay, A. S., Terbush, B., … Vinci, M. (2020). Developing post-alert messaging for ShakeAlert, the earthquake early warning system for the West Coast of the United States of America. *International Journal of Disaster Risk Reduction*, *50*(May), 101713. https://doi.org/10.1016/j.ijdrr.2020.101713

McBride, S. K., Llenos, A. L., Page, M. T., & Van Der Elst, N. (2019). #earthquakeAdvisory: Exploring discourse between government officials, news media, and social media during the 2016 Bombay beach swarm. *Seismological Research Letters*, *91*(1), 438–451. https://doi.org/10.1785/0220190082

McBride, S. K., Michael, A., Wein, A. M., Hardebeck, J., Becker, J. S., Potter, S. H., … Van Der Elst, N. (2018). Developing earthquake forecast templates for fast and effective communication. *11th National Conference on Earthquake Engineering 2018, NCEE 2018: Integrating Science, Engineering, and Policy*, *12*(September 2019), 7310–7316.

Menge, D. N. L., MacPherson, A. C., Bytnerowicz, T. A., Quebbeman, A. W., Schwartz, N. B., Taylor, B. N., & Wolf, A. A. (2018). Logarithmic scales in ecological data presentation may cause misinterpretation. *Nature Ecology and Evolution*, *2*(9), 1393–1402. https://doi.org/10.1038/s41559-018-0610-7

Michael, A. J., McBride, S. K., Hardebeck, J. L., Barall, M., Martinez, E., Page, M. T., … Wein, A. M. (2019). Statistical seismology and communication of the USGS operational aftershock forecasts for the 30 November 2018 Mw 7.1 Anchorage, Alaska, earthquake. *Seismological Research Letters*, *91*(1), 153–173. https://doi.org/10.1785/0220190196

Morss, R. E., Demuth, J. L., Lazo, J. K., Morss, R. E., Demuth, J. L., & Lazo, J. K. (2008). Communicating Uncertainty in Weather Forecasts: A Survey of the U.S. Public. *Weather and Forecasting*, *23*(5), 974–991. https://doi.org/10.1175/2008WAF2007088.1

Nakayachi, K., Becker, J. S., Potter, S. H., & Dixon, M. (2019). Residents' Reactions to Earthquake Early Warnings in Japan. *Risk Analysis*, *39*(8), 1723–1740. https://doi.org/10.1111/risa.13306

Peresan, A., Kossobokov, V. G., & Panza, G. F. (2012). Operational earthquake forecast/prediction. *Rendiconti Lincei*, *23*(2), 131–138. https://doi.org/10.1007/s12210-012-0171-7

Perry, S. C., Blanpied, M. L., Burkett, E. R., Campbell, N. M., Carlson, A., Cox, D. A., … Zarcadoolas, C. (2016). *Get your science used—Six guidelines to improve your products*. *Circular*. Reston, VA. https://doi.org/10.3133/cir1419

Pighin, S., Savadori, L., Barilli, E., Cremonesi, L., Ferrari, M., & Bonnefon, J. F. (2011). The 1-in-X effect on the subjective assessment of medical probabilities. *Medical Decision Making*, *31*(5), 721–729. https://doi.org/10.1177/0272989X11403490

Romano, A., Sotis, C., Dominioni, G., & Guidi, S. (2020). The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences. *Health Economics (United Kingdom)*, *29*(11), 1482–1494. https://doi.org/10.1002/hec.4143

Roth, E., Morgan, M. G., Fischhoff, B., Lave, L., & Bostrom, A. (1990). What do we know about making risk comparisons? *Risk Analysis*, *10*(3), 375–387. https://doi.org/10.1111/j.1539-6924.1990.tb00520.x

Sandman, P. M., Weinstein, N. D., & Miller, P. (1994a). High risk or low: how location on a "risk ladder" affects perceived risk. *Risk Analysis : An Official Publication of the Society for Risk Analysis*, *14*(1), 35–45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8146401

Sandman, P. M., Weinstein, N. D., & Miller, P. (1994b). High Risk or Low: How Location on a "Risk Ladder" Affects Perceived Risk. *Risk Analysis*, *14*(1), 35–45. https://doi.org/10.1111/j.1539-6924.1994.tb00026.x

Slovic, P., Fischoff, B., & Lichtenstein, S. (1981). Perceived Risk: Psychological Factors and Social Implications. *Proceedings of the Royal Society of London A*, *376*, 17–34. https://doi.org/10.1017/CBO9781107415324.004

Slovic, P., Kraus, N., & Covello, V. T. (1990). What Should We Know About Making Risk Comparisons? *Risk Analysis*, *10*(3), 389–392. https://doi.org/10.1111/j.1539-6924.1990.tb00521.x

Spiegelhalter, D. J. (2017). Risk and Uncertainty Communication. *Annual Review of Statistics and Its Application*, *4*(1), 31–60. https://doi.org/10.1146/annurev-statistics-010814-020148

Stallings, S. P., & Paling, J. E. (2001). New tool for presenting risk in obstetrics and gynecology. *Obstetrics and Gynecology*, *98*(2), 345–349. https://doi.org/10.1016/S0029-7844(00)01182-0

Sutton, J., Fischer, L., James, L. E., & Sheff, S. E. (2020). Earthquake early warning message testing: Visual attention, behavioral responses, and message perceptions. *International Journal of Disaster Risk Reduction*, *49*(March), 101664. https://doi.org/10.1016/j.ijdrr.2020.101664

Tan, M. L., Prasanna, R., Becker, J. S., Brown, A., Kenney, C., Lambie, E., … Alwis, D. De. (2021). Outlook for earthquake early warning for Aotearoa New Zealand : Insights from initiating a community-of-practice. *2021 Annual Technical Conference for the New Zealand Society for Earthquake Engineering*, 55–63.

Thompson, M. A., Lindsay, J. M., & Gaillard, J. C. (2015). The influence of probabilistic volcanic hazard map properties on hazard communication. *Journal of Applied Volcanology*, *4*(1).

https://doi.org/10.1186/s13617-015-0023-0
van der Bles, A. M., van der Linden, S., Freeman, A. L. J. J., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(14), 7672–7683. https://doi.org/10.1073/pnas.1913678117
Wang, K., & Rogers, G. C. (2014). Earthquake preparedness should not fluctuate on a daily or weekly basis. *Seismological Research Letters*, *85*(3), 569–571. https://doi.org/10.1785/0220130195
Wein, A., Potter, S., Johal, S., Doyle, E., & Becker, J. (2016). Communicating with the public during an earthquake sequence: Improving communication of geoscience. *Seismological Research Letters*, *87*(1), 112–118. https://doi.org/10.1785/0220150113
Woo, G., & Marzocchi, W. (2014). Operational Earthquake Forecasting and Decision-Making, 353–367. https://doi.org/10.1007/978-3-642-12233-0_18
Yamagishi, K. (1997). When a 12.86 % Mortality is More Dangerous than 24.14 %: Implications for Risk Communication. *Applied Cognitive Psychology*, *11*(January), 495–506. https://doi.org/10.1002/(SICI)1099-0720(199712)11:6<495::AID-ACP481>3.0.CO;2-J