The Communication of Uncertain Scientific Advice During Natural Hazard Events

Emma E. H. Doyle, *Massey University*

David M. Johnston, *GNS Science, Lower Hutt, NZ*

John McClure, *Victoria University of Wellington*

Douglas Paton, *University of Tasmania*

During natural hazard crises such as earthquakes, tsunami, and volcanic eruptions, a number of critical challenges arise in emergency management decision-making. A multidisciplinary approach bridging psychology and natural hazard sciences has the potential to enhance the quality of these decisions. Psychological research into the public understanding of different phrasings of probability has identified that the framing, directionality and probabilistic format can influence people's understanding, affecting their action choices. We present results identifying that translations of verbal to numerical probability phrases differ between scientists and non-scientists, and that translation tables such as those used for the International Panel on Climate Change reports should be developed for natural hazards. In addition we present a preliminary result illustrating that individuals may 'shift' the likelihood of an event towards the end of a time window.

New Zealand is a country at risk from numerous extreme natural hazards that pose a threat to life, infrastructure and business. These include explosive volcanic eruptions, earthquakes and tsunami. Emergency management of these events involves a number of critical and challenging decisions often based on limited and uncertain information, incorporating an integration of the wide range of scientific opinions, model outputs, and outcome scenarios. The challenges inherent in this process were evident in the response and recovery management phases of the September 2010 and February 2011 Canterbury earthquakes.

These complex issues can arise in the management of volcanic crises, which, during the lead up to a potential eruption and the management of the ensuing volcanic crisis, present considerable uncertainty to emergency management decision makers.

This paper includes an introduction to the 'volcano problem', followed by a review of emergency management in New Zealand, and of Exercise Ruaumoko, a simulation which tested the use of a scientific advisory group during the lead up to an imaginary eruption in Auckland. We then summarise the literature on the communication of verbal and numerical probabilities, with a discussion of the translation table approach that the International Panel on Climate Change (IPCC) adopts. Finally, we present some preliminary results of a survey to assess the differences between scientists' and non-scientists' translations of verbal probability phrases to numerical equivalents, and their perceptions of event likelihoods across time windows for multi-day statements.

An Introduction to the Volcano Problem

Prior to a volcanic eruption, many volcanoes exhibit precursory signals that indicate an eruption may occur. These can range from an increase in volcanic type earthquakes that may be felt by the local community or detected through sensitive seismometers, to changes in steam or other geothermal emissions from the volcano, and deformation of the volcano itself due to the inject of magma beneath the surface (see Johnston et al., 2002, for a review). However, these precursory signals are only hints that something *may* be happening. The lead up period to a volcanic eruption can range from hours (e.g., 19 hours at Rabaul, Papa New Guinea, see Blong & McKee, 1995), to many months (e.g., 11 months at Mt Pinatubo, Philippines, see Newhall & Punongbayan, 1996b), and may not result in an eruption at all (e.g., Mammoth Mt/Long Valley Caldera, USA in 1980s, see Hill, 1998). In addition, when an eruption does occur there can be much uncertainty about the size of the eruption, and the type of impacts that may result. Thus, volcanic eruptions create an extremely uncertain environment for emergency management planning and the information and decision management required for effective response (Paton & Auld, 2006; Paton, Johnston, & Houghton, 1998), as critical decision makers balance the issue of life safety and community continuity through the crises. Added to the uncertainty implicit in managing the event itself, uncertainty emerges in relation to activities such as deciding on and advising of the need for evacuation in the context of concerns about making an "economically disastrous, unnecessary evacuation" (Tazieff, 1983, as cited in Woo, 2008, p. 88).

From a volcanological view, the successful management and response to the lead up to an eruption is thus fundamentally dependent upon:

(a) the geological knowledge, and the enhancement of this knowledge through the continued monitoring of the volcano (see reviews in Sparks, 2003; Tilling, 2008);

(b) the communication between the scientific advisors and the emergency management community to guide their critical decisions both before (reduction, readiness), during (response), and after (recovery) a crisis (see review in Doyle & Johnston, 2011) ; and

(c) the onward communication of this advice to the public through public education programmes and warnings (Leonard et al., 2008).

The focus of the research reported in this paper is to explore the link between scientific advisory groups and the emergency management community, and how uncertainty impacts this communication. At steps b and c, it is also important to understand how agencies and community members interpret and use information and to accommodate the fact that the mental models of the latter can differ from each other and from the scientists producing the data. Thus, irrespective of the objective quality of the information made available by the scientific community, its ability to have the desired effect is influenced by how it is interpreted and filtered as it is transmitted to various recipients. A good example of the many layers of this interpretation is that represented by the multi-tiered nature of emergency management organisations, as explained in the next section.

Emergency Management in New Zealand

In New Zealand, civil defence and emergency management is coordinated through a three-tiered structure: national, regional, and local council/territorial authority (Lee, 2010). At the national level, the Ministry of Civil Defence and Emergency Management (MCDEM) promotes and manages policies and programmes for civil defence and emergency management (MCDEM, 2008a). During a national crisis, MCDEM will lead the response via the National Crisis Management Centre (NCMC), which is a national level Emergency Operations Centre (EOC). An EOC such as this is a facility for central command and control, which when activated during a response is responsible for carrying out disaster management functions (see NZ Fire Services Commission, 1998). Through this process the response of multiple agencies is handled (fire, police protective agencies, Civil Defence, volunteers, etc). The NCMC liaises with and supports the 16 regional council CDEM groups across New Zealand, each of which operates their own Group EOC (GEOC) and in turn coordinates and supports EOCs at the local council level (Lee, 2010).

There are a number of key strategic positions within an EOC, organised around the NZ Coordinated Incident Management System (CIMS, NZ Fire Services Commission, 1998, p. 14). The four main components are:

Control: management of the incident;

Planning and Intelligence: collection and analysis of incident information and planning of response activities;

Operations: direction of an agency's resources in combating the incident; and

Logistics: provision of facilities, services and materials required to combat the incident.

This CIMS structure enables personnel from different agencies, police, fire, and beyond, to work directly with their equivalent counterpart in another agency. The majority of the scientific and geological advice is thus directly communicated to the Planning and Intelligence desk, where it is utilised in the generation of situation reports and action plans. In addition, it is often also communicated directly to the Controller managing the incident, and through additional pathways to the wider CDEM community and the general public via bulletins, broadcasts and warnings (e.g., GeoNet daily volcanic bulletins during a crisis, the Pacific Tsunami Warning Centre alerts, and MetService severe weather forecasts). Crucial to realising the response benefits of CIMS training are exercises and simulations that can identify interpretation problems, allow their rectification and develop people's capacities for decision making under stress (Paton & Flin, 1999), with developing capacity to understand and use science advice being a key goal of these exercises.

Learning from exercises: The role of science advisors

Communication problems have occurred in numerous volcanic hazard crises due to conflicting scientific advice either from internal and external agencies, or due to the presence of a wide range of scientific advisory bodies and individuals. Thus, experience from previous volcanic crises has led to the practice of communicating scientific advice from one source during a volcanic crisis (see review in Doyle & Johnston, 2011, and the International Association for Chemistry and the Earth's Interior (IAVCEI) Subcomittee for Crisis Protocols, 1999). In NZ, this has been undertaken through the development of Scientific Advisory Groups (SAGs) established to bring the advice from various scientific agencies together. There are many different natural hazard Science Advisory Groups within NZ, including the Central Plateau Volcanic Advisory Group (CPVAG) to advise officials about the Central Volcanoes of the North Island, the Auckland Volcanic Scientific Advisory Group (AVSAG) to advise officials about the volcanic field under Auckland, and the Tsunami Expert Panel (TEP) which forms in response to a local, regional, or distant source earthquake.

The process of the AVSAG advice provision was tested out from November 2007 to March 2008 through Exercise Ruaumoko, which was run as part of MCDEM's National Exercise Programme. Through a representative governance group, MCDEM and the 16 regional council CDEM groups manage this ongoing national programme to encourage the practicing and continuous improvement of response planning, as well as the building of interagency relationships and processes (MCDEM, 2009). These exercises range from Tier 1 (Local Exercise run by an individual organisation) to Tier 4 (National Exercise including central government). Exercise Ruaumoko was a Tier 4 level exercise, and was run to test the local, regional, and national arrangements for dealing with the impact of a large natural hazard event on a major population centre (MCDEM, 2008b).

Auckland was chosen as it sits on a 'monogenetic' basalt volcanic field (Auckland Volcanic Field, AVF), where individual eruptions can occur at different distributed volcanic vents, with more than 49 volcanic centres identified in the 360km field so far. The largest and youngest eruption occurred approximately 600 years ago, forming Rangitoto Island (see review in Lindsay et al., 2009). For the AVF, precursory lead times between detectable eruption precursors and an eruption at the surface can range from months, to weeks, to less than a few days (see Blake, Wilson, Smith, & Leonard, 2006), or may not lead to an eruption at all as magma 'stalls' en route to the surface leading to what may be considered to be a 'failed eruption'. As eruptions can occur anywhere within the AVF, and the location may not be known until magma is very close to the surface, emergency management decisions will be typified by a high degree of uncertainty due to the eruption timing, location, severity, hazards, impacts and consequences (Lindsay et al., 2009; MCDEM, 2008b).

The scenario in Exercise Ruaumoko focused on the lead-up to a volcanic eruption in the Auckland metropolitan area, and the exercise was the first full test of the AVSAG advisory process (see reviews in MCDEM, 2008b; McDowell, 2008; Smith, 2009). This advisory group represented a wide range of expertise including members from universities, Crown Research Institutes, consultancies, and members of local and national CDEM groups. Advice was delivered during the 'event' through a tripartite sub-group system (Monitoring, Volcanology, and Social) all of which reported upwards to a smaller core SAG. This SAG then liaised directly with the NCMC and the Auckland Group EOC through teleconferences and two on-site liaison officers, who acted as a further information conduit between AVSAG, GeoNet (the monitoring arm of GNS Science), and the CDEM sector.

A number of reviews were conducted after Exercise Ruaumoko, both at the National level (MCDEM, 2008b) and at the Auckland Regional Level (McDowell, 2008), identifying that the structure of science advice resulted in it being well delivered, clear, timely and very valuable. The use of on-site liaison officers was found to be very beneficial, enabling further translation and use of the expert advice by the emergency managers in the NCMC and the Auckland GEOC. A recommendation was the demonstrated importance of having scientific advice provided by "one trusted source" through AVSAG, as it helped to prevent conflicting or confusing messages (MCDEM, 2008b). Suggested improvements included adjustments to the finer details of the advisory process and structure to encourage more integration between the different sub-groups of the AVSAG, to prevent a disconnect between local and national advice provision to the Auckland GEOC and NCMC, and to ensure that the science advice and science research response, capability, and process, remain integrated (Cronin, 2008; MCDEM, 2008b; McDowell, 2008; Smith, 2009). We will not discuss these further here, except to say that the advisory group model is still undergoing development (Smith, 2009) and will no doubt evolve further to encompass lessons learnt from many recent hazard events and exercises in New Zealand (including the September 2010 Canterbury and February 2011 Christchurch earthquakes, the Pike River Mine disaster 2010, and the Tauranga oil spill 2011).

Communicating Uncertainty and the Use of Probabilities

Communicating from 'one trusted source' does not imply that the communication should be a consensus opinion, or that the communication does not include information about the associated uncertainty in the knowledge, data, or outcome, and thus it is important to identify how best to communicate these aspects.

There is much discourse in the psychological literature as to whether revealing the uncertainties associated with a risk assessment will strengthen or decrease trust in a risk assessor and their message (see reviews in Miles & Frewer, 2003; Wiedemann, Borner, & Schultz, 2008). On the one hand, the communication of uncertainty has been suggested to enhance credibility and trustworthiness of the information provider. On the other, however, studies have suggested that it can decrease people's trust and the credibility of the provider. It has also been suggested that the provision of uncertainty can allow people to justify inaction or their own agenda, or to perceive the risk as being higher or lower than it actually is depending on their personal attitudes.

To address the many risks and uncertainties involved in volcanic eruptions, due to their complex nature, it has become increasingly popular for scientists to use probability statements in their communications. These probabilistic forecasts usually involve knowledge of both the dynamical phenomena and the uncertainties involved (Sparks, 2003). Recently, there has been a move to include pre-defined thresholds of probability based on a cost benefit analysis, prompted by a desire to make objective decisions via quantitative volcanic risk metrics (Lindsay et al., 2009; Woo, 2008). These cost-benefit analysis tools, and the use of forecasting systems such as Bayesian Event Trees for eruptions (Aspinall & Cooke, 1998; Marzocchi & Woo, 2007) are viewed as being highly advantageous for the decision-making process of the scientists, as it clarifies decision thresholds as well as optimising the decision-making time, as well as offering the hindsight ability to clearly explain how a decision was made (Lindsay et al., 2009).

However, Haynes, Barclay, and Pidgeon (2008, p. 263) found at Montserrat Volcano Observatory, West Indies, that the use of probabilities "was considered to complicate communications as the likelihoods and associated uncertainties were neither well-explained nor understood". In addition, Cronin (2008) recognised, in a review of Exercise Ruaumoko, a need for the identification of protocols for communicating probabilities and uncertainties during volcanic crises to avoid misinterpretations during forecast communications. The IAVCEI Subcommittee For Crisis Protocols (1999, p. 330) recommend the use of "probabilities to calibrate qualitative assessments of risk". Other volcanic crisis communication guidelines (e.g., McGuire, Solana, Kilburn, & Sanderson, 2009, p. 67) recommend that "qualitative, non-technical statements yield more positive reactions among non-scientists". In particular, these authors highlight that confusion can occur due to "a limited public understanding of … concepts such as probabilities in the forecasts", and recommend that "percentages or proportions should be used carefully and sparingly and backed up by a more general statement" (p. 68).

An overview of lessons from the literature on communicating uncertainty

In Exercise Ruaumoko a number of probabilistic statements were included in both the daily GeoNet volcanic bulletins, and the AVSAG communications, for example:

- … *"If magma ascent continuous [sic] at the present rate an eruption is likely in the next 2-3 days."* (Exercise Ruaumoko Science Alert Bulletin, AK-08/09, 11 March 2008)
- … *"Within this zone there is a 25-50% probability of an eruption within the next 24 hours increasing to 75-90% within the next 48 hours."* … (Exercise Ruaumoko Science Alert Bulletin, AK-08/13, 12 March 2008)

Looking at the first statement, an immediate question arises as to what "likely" actually means to the emergency managers. The emergency managers may interpret the likelihood quite differently to that intended by the scientists, and thus make disproportionate action choices. In the second statement, questions arise as to whether the numerical probabilities are interpreted by the emergency managers as high or low risk prompting either action, or inaction, and how this compares to the scientists' understanding. Anecdotal discussions with participants after Exercise Ruaumoko raised the issue that the language with which the forecasts were communicated was being understood differently between the scientists and the emergency managers, whereby one would see 50% chance as being 'low' and another as it being 'high' and requiring immediate action. These questions require consideration both in the context of lessons learnt from the literature (discussed next), and through further direct investigations for the volcanic risk communication problem (discussed later).

Communicating verbal and numerical probabilities

The communication of probabilistic statements has been studied extensively in the literature, and a number of lessons can be drawn from this for the communication of probabilistic forecasts during natural hazard events. These statements, whether they are in a numeric or linguistic format, can commonly be misinterpreted because their framing, directionality and probabilistic format can bias people's understanding, thereby affecting their action choices (e.g., Budescu, Broomell, & Por, 2009; Honda & Yamagishi, 2006; Joslyn, Nadav-Greenberg, Taing, & Nichols, 2009; Karelitz & Budescu, 2004; Lipkus, 2010; Teigen & Brun, 1999). Verbal and linguistic probabilities include phrases such as *unlikely*, *likely*, *certain*, *uncertain* (see Risbey & Kandlikar, 2007; Teigen & Brun, 1999), with modifiers such as *virtually*, *very*, *exceptionally*, *extremely* (see Budescu et al., 2009; Dhami & Wallsten, 2005; Lipkus, 2010; Teigen & Brun, 1999). Experiments conducted by Brun and Teigen (1988) demonstrated that the term 'likely' can be translated to a numerical probability of $p = 0.67$ with a standard deviation of 0.16, and this mean value can change to 0.71 or 0.59 depending on the experimental context. Thus, one person may view 'likely' to represent a probability as low as 51% and another as high as 83% (see also Lipkus, 2010).

In addition to the translation issue discussed above, Teigen and Brun (1999) identified that semantic issues can also cause miscommunications. These occur when the verbal phrases convey additional information beyond that which would be communicated via their numerical equivalents, as described by their directionality (Budescu, Karelitz, & Wallsten, 2003; Honda & Yamagishi, 2006; Joslyn & Nichols, 2009; Teigen & Brun, 1999), or the framing of the outcome (Kuhberger, 1998; Levin, Schneider, & Gaeth, 1998). The context and outcome severity of the occurrence has also been found to affect people's likelihood perceptions. Studies have demonstrated that people can view a probability as being greater than it actually is if the severity of the outcome is high (e.g., Bruine De Bruin, Fischhoff, Millstein, & Halpern-Felsher, 2000; Patt & Dessai, 2005). Thus, people will interpret a ‘slight chance of cancer’ as being of greater likelihood than a ‘slight chance of a sprained ankle’ (Weber, 1994; Windschitl & Weber, 1999).

Numerical, or frequentist, probabilistic statements have also been found to be subjected to the same affects (e.g., Bruine De Bruin et al., 2000; Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Joslyn & Nichols, 2009). For example, Gigerenzer and Edwards (2003) state that there are three types of numerical representations that can cause confusion: single event probabilities, conditional probabilities, and relative risks. This confusion arises because it can be difficult to understand the class of events a probability or percentage is referring to. For example, a single event probability such as “a 30% chance of rain tomorrow” can cause misunderstanding as it does not specify the class of events and thus some could interpret this as 30% of the area, or 30% of the time, or 30% of days like tomorrow (Gigerenzer, Hertwig, Broek, Fasolo, & Katsikopoulos, 2005).

Using translation tables to communicate probabilities

Miscommunication of verbal probabilities between experts and non-experts has been investigated in a number of fields, including medical practitioners and the general public (Brun & Teigen, 1988), as well as climate scientists and policy makers (Patt & Dessai, 2005). Patt and Dessai (2005) highlight the importance of considering your target audience when communicating an uncertainty, suggesting for example that the IPCC reports use a pluralistic approach with highly sophisticated parts of the report using a numeric format, and the more general chapters using verbal phrases and narratives.

However, even though there is a variance in people’s numerical interpretation, verbal probability phrases are generally better understood than their numerical counterparts (Patt & Schrag, 2003; Wallsten, Fillenbaum & Cox, 1986) and are thus still the preferred form of communication in many fields.

In some fields, there has been a move to formalise the translation of verbal probability phrases. For example, since 2002 the IPCC reports have utilized qualitative descriptors for probability, as illustrated in Table 1.

▼ **Table 1**: IPCC Qualitative Descriptors used for the Third Assessment Report Climate Change 2001, as given in Patt & Schrag (2003).

Probability range	Descriptive term
<1%	Extremely unlikely
1-10%	Very unlikely
10-33%	Unlikely
33-66%	Medium likelihood
66-90%	Likely
90-99%	Very likely
>99%	Virtually certain

This process was initiated for the Third Assessment Report Climate Change 2001 (Houghton et al., 2002; herein referred to as IPCC3), in response to the recommendation of Moss and Schneider (2000) that the IPCC lead authors should communicate uncertainty via a seven-step approach (see reviews in Patt & Schrag, 2003; Risbey & Kandlikar, 2007). However, as discussed by Karelitz and Budescu (2004, p. 26), a “drawback of standardised verbal scales is the difficulty of most people to suppress the meanings they normally associate with these terms”.

Patt and Dessai (2005) caution that when defining probability words and phrases, one should explain that such a rigid framework does not necessarily match people’s intuitive use of the language, in the hope that this will prevent bias in conscientious readers. Budescu et al. (2009) have additionally found that the verbal probabilities in the 2007 IPCC report (herein referred to as IPCC4) may have implied higher levels of imprecision than are actually present. To address this, they recommend that an alternative form of communication should be used, where both verbal and numerical terms are used together, with the inclusion of a range for the numerical values where the range matches the uncertainty of the target events.

A Survey on the Communication of Probabilities in Volcanic Crises

As discussed above, during a volcanic crisis event or exercise, a multitude of verbal and numerical probabilistic statements can be produced on an almost daily basis. These statements often form the fundamental basis of the decisions made by emergency management personnel in their response to the crisis, and thus it is vital that the potential for miscommunication and misunderstanding is reduced as much as possible.

Based on findings from the judgment literature research community, and the fact that scientists in volcanic crises are currently using deterministic, verbal, numerical, and time window predictive statements, there is a need now to identify differences in the scientists’ and civil authorities’ perceptions of the language used in these communications.

To address this, we conducted three experiments via an online survey tool, to investigate:

- The differences in translations between verbal and numerical probability phrases.
- The perception of likelihood distributions within time windows.
- The relationship between the perception of these distributions and action choice scenarios (in the manner of Joslyn et al., 2009).

Survey method

The multi-part online survey tool featured both within- and between-subject design and was administered through the Qualtrics Survey Research Suite software (Qualtrics Labs Inc., Provo, UT, USA, Version 2.03s, Copyright ©2011). This enabled the randomisation of questions within parts 1, 2 and 3, as well as the random allocation of participants to either Experiment Group A or B for parts 2 and 3. In part 2, each experiment group contained questions with either verbal or numerical phrases using the translations outlined by the IPCC3 (Table 1), while in part 3 each experiment group contained statements that utilised either the phrasing "*in*" or "*within*" to describe time windows, as these were used interchangeably during Exercise Ruaumoko.

Participants were recruited from scientists in the natural hazard community of New Zealand (e.g., GNS Science, NIWA), from both physical and social scientists across NZ universities, and from civil authorities across the nation (e.g. MCDEM, CDEM, emergency and protective services, lifelines, etc.). In addition, the survey was delivered internationally to capture both the NZ and global perspectives, of importance due to the internationalisation of both the volcanological and emergency management sectors.

Participants were directed to the online survey tool through a snowball approach via email contact with individuals in each organisation, and through advertisement in bulletins and on on-line notice boards, such as the MCDEM e-bulletin, the international 'Volcano Listserv' (run by Arizona State University), the bulletin board of the Comprehensive Emergency Management Research Network (CEMR), and in the *Oceania* newsletter of the International Association of Emergency Managers.

The survey was anonymous, and participants were asked to identify their primary employment sector, including specific options for:

- Scientific or technical (agency, university or research institute);
- Central/national government, civil defence, emergency Management (*Ministry, agency, etc*);
- Local/regional government, civil defence, emergency management (*Council, agency, etc*);
- Public safety, emergency services (police, fire, ambulance, rescue, response, etc);
- Lifelines (infrastructure, water, telecommunications, electricity, transportation, gas, etc);
- Other.

From here on in this study we refer to category 1 as *scientists*, and categories 2 to 6 as *non-scientists*. This definition is based upon the multi-disciplinary nature of both the Scientific Advisory Groups (which incorporates geology, social science, economics), and the emergency management community (which incorporates lifeline management, CDEM, defence, fire, police, etc). Additional background questions included educational background, geographical region of residence, employer name, job role, and gender. In total, there were 179 participants who completed the survey, with 92 identifying as scientists, 85 as non-scientists, and 2 unidentified, and 47 choosing to identify their gender as women, 90 as men. We briefly report here on some initial results from part 1 of the survey tool, and an example question from part 2.

Preliminary results: translating verbal to numerical terms

The aim of part 1 of the survey was to explore the translation of vague verbal probabilistic terms, such as the term 'likely' used in the example Ruaumoko statement discussed above. The terms '*extremely unlikely*', '*very unlikely*', '*unlikely*', '*medium likelihood*', '*likely*', '*very likely*', '*virtually certain*' were all examined, to investigate how the translation of these terms compares to the guidelines outlined in the IPCC3 (Table 1).

Initially, participants were shown each of these phrases in randomised, context free statements; these were then followed by four randomised context statements such as '*At the current magma ascent rate, an eruption is likely*'.

All participants received the same statements, and following the methodology of Budescu et al. (2009), each participant was asked to rate on a numerical sliding bar scale 'Your BEST estimate' of the probability conveyed, as well as 'THE LOWEST possible' and 'THE HIGHEST possible' probabilities. Figure 1 shows the online display. We report below the preliminary results from the context free statements.

▼**Figure 1**: A screen shot of the online survey format for part 1, which assessed participant's translations from verbal to numerical probabilities using the verbal terms in the IPCC3 report (Table 1).

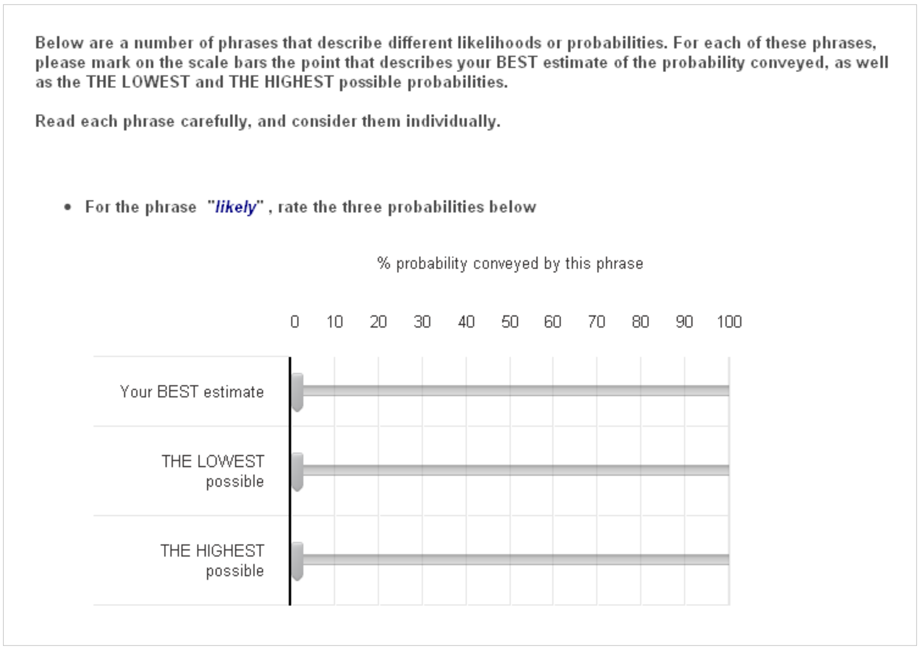


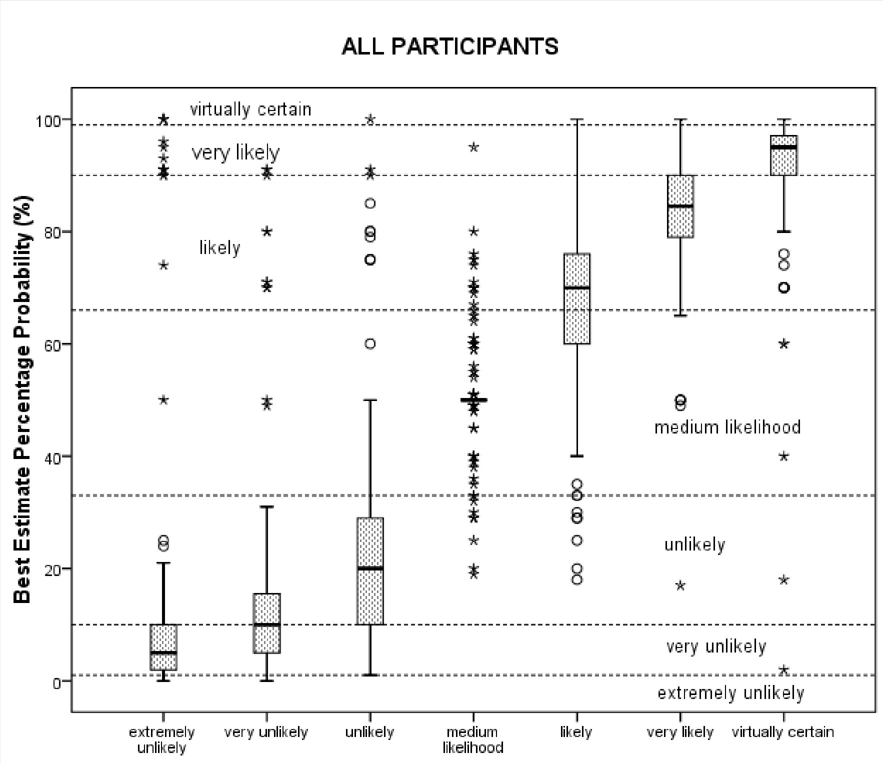
Figure 2 illustrates the 'BEST estimate' translation for all participants, Figure 3 shows the translation for the group that identified themselves as *scientists*, and Figure 4 illustrates the translation for the group that identified themselves within the categories of *non-scientists*.

Initial impressions from these figures is that the BEST estimates identified by the participants do not conform well with the IPCC3 guidelines at the extremes, and overall performance is worse for the more positive terms (> *medium likelihood*) with observable differences between *scientists* and *non-scientists*. The BEST estimates of medium likelihood are in a very narrow and extremely consistent range across both groups, suggesting that the category in the IPCC may be too wide.

Following a method similar to that of Budescu et al. (2009), we identified whether the 'LOWEST possible' and 'HIGHEST possible' probabilities chosen by participants were *consistent* with the IPCC3 guidelines. We refer to these two chosen values as the *'RANGE estimate'*), which is deemed *consistent* if both the upper value and the lower value are within the range outlined in Table 1, and as *inconsistent* if they are outside the guideline range, and *partially consistent* otherwise. The same approach was also adopted for the 'BEST estimate', but using only the categories *consistent* and *inconsistent.*

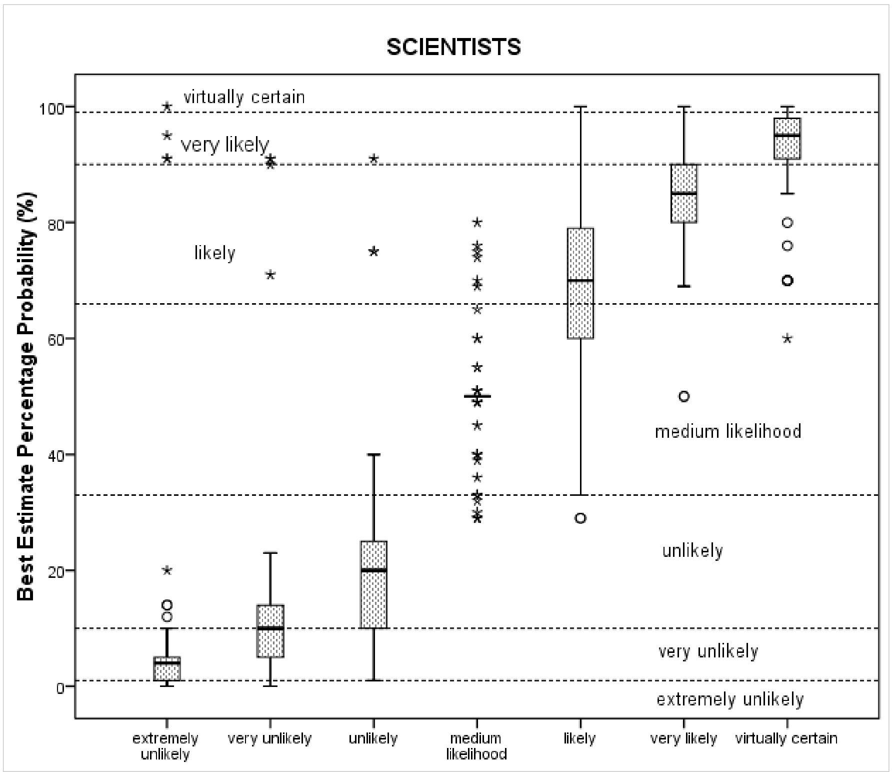
For the calculation of consistent, partially consistent, and inconsistent, we use the IPCC3 translation table given in Patt and & Schrag (2003). This differs from Budescu et al. (2009), who use the IPCC4 tables.

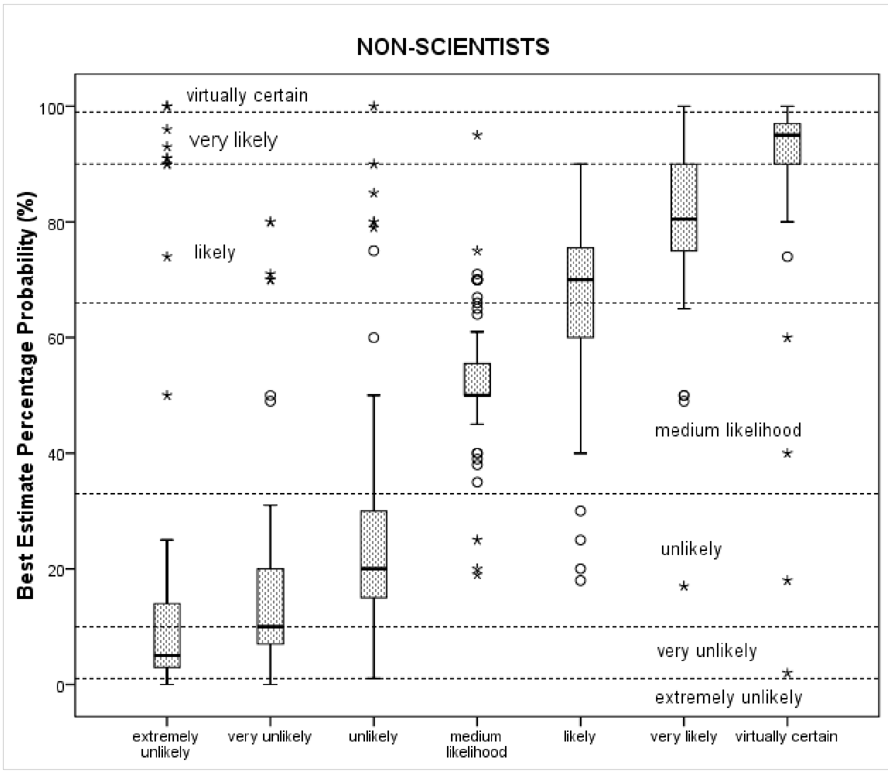
We use the IPCC3 table because the probability ranges are bounded (e.g., unlikely corresponds to 10-33%), whereas in the IPCC4 table the probability ranges are unbounded (e.g., unlikely corresponds to <33%). The former approach is more suited to volcanological communications, and IPCC reports give no explanation as to why the translation table was changed.



▲**Figure 2**: The central 50% of numerical translations (boxes) of each verbal probability term, for all participants that took the survey. The solid lines within the box represent the median, and the whiskers represent 1.5 times the inter-quartile range above the first quartile and below the third. Circles indicate outliers, and stars indicate extremes. The horizontal dashed lines represent the translation boundaries given in the IPCC3 report (Table 1), as also indicated by text in the figure.

▼**Figure 3**: The central 50% of numerical translations (boxes) of each verbal probability term, for all participants that identified as *scientists;* key as for Figure 2





▶Figure 4: The central 50% of numerical translations (boxes) of each verbal probability term, for all participants that identified as *non-scientists*. Key as for Figure 2

Tables 2 and 3 summarise these results for a) all participants, b) scientists, and c) non-scientists. As also found by Budescu et al. (2009), consistency with the IPCC guidelines was low, 'especially for phrases that convey more extreme probabilities' (ibid, p. 302). We also find that there is a significant difference between scientists and non-scientists for the *BEST estimate* of the term *very unlikely*, χ^2 (1, N=167) = 5.483, *p*<0.05. For the *RANGE estimate* a significant difference was found between scientists and non-scientists for the terms *unlikely*, χ^2(2, N=148) = 7.3, *p*<0.05, and *likely*, χ^2(2,N=169) = 11.693, p<0.05. We also find for this *RANGE estimate* that a significant difference exists between the values chosen by scientists and non-scientists for all terms, χ^2(1,N=168) = 5.017, p<0.05, when the term *medium likelihood* is excluded and the categories *consistent* and *partially consistent* are combined to avoid low expected frequencies due to low consistency for extreme terms.

▶Table 2: The Percentage of Participants Whose 'RANGE Estimate' (Defined by the HIGHEST and LOWEST Possible Probability Values Chosen by Participants) are Consistent (C), Partially Consistent (PC), or Inconsistent (I) with the IPCC3 Range Boundaries described in Table 1

	All (n=179)			Scientists (n=92)			Non-Scientists (n=85)		
Phrase	**C**	**PC**	**I**	**C**	**PC**	**I**	**C**	**PC**	**I**
Extremely unlikely	3.5	47.5	48.9	2.6	53.8	43.6	4.8	39.7	55.6
Very unlikely	6.3	83.3	10.4	6.4	79.5	14.1	6.1	87.9	6.1
Unlikely	18.9	61.5	19.6	17.6	54.1	28.4	20.3	68.9	10.8
Medium likelihood	39.3	41.1	19.6	39.1	35.6	25.3	39.5	46.9	13.6
Likely	10.1	61.5	28.4	5.7	55.2	39.1	14.6	68.3	17.1
Very likely	1.2	79	19.8	1.2	74.4	24.4	1.2	84	14.8
Virtually certain	1.8	58.4	39.8	0	63.5	36.5	3.7	53.1	43.2
All terms	11.6	61.8	26.6	10.4	59.4	30.2	12.9	64.1	23
All terms except Medium likelihood	15.1	65.3	19.6	14	59.8	26.3	16.3	71.2	12.5

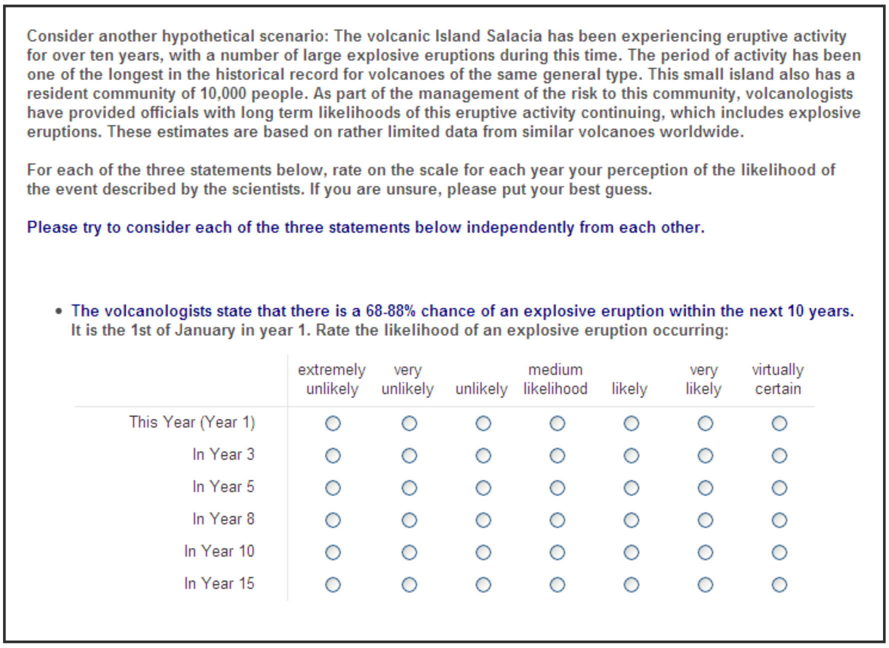
▶Table 3: The Percentage of Participants Whose BEST Estimates are Consistent or Inconsistent with the IPCC3 Range Boundaries described in Table 1

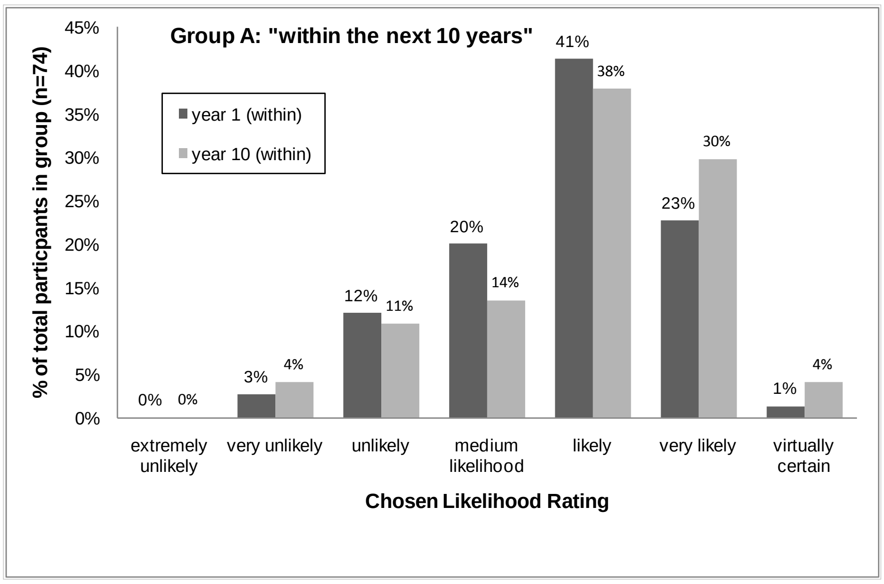
	All (n=179)		Scientists (n=92)		Non-Scientists (n=85)	
Phrase	**C**	**I**	**C**	**I**	**C**	**I**
Extremely unlikely	6.7	93.3	8.2	91.8	5.1	94.9
Very unlikely	63.5	36.5	72.4	27.6	53.8	46.3
Unlikely	71.3	28.7	74.7	25.3	67.5	32.5
Medium likelihood	87.1	12.9	87.5	12.5	86.7	13.3
Likely	63.6	36.4	64	36	63.1	36.9
Very likely	30.6	69.4	34.1	65.9	26.8	73.2
Virtually certain	4.7	95.3	2.3	97.7	7.3	92.7
All terms	42.1	57.9	44.9	55.1	39	61
All terms except Medium likelihood	22.2	77.8	24.7	75.3	19.5	80.5

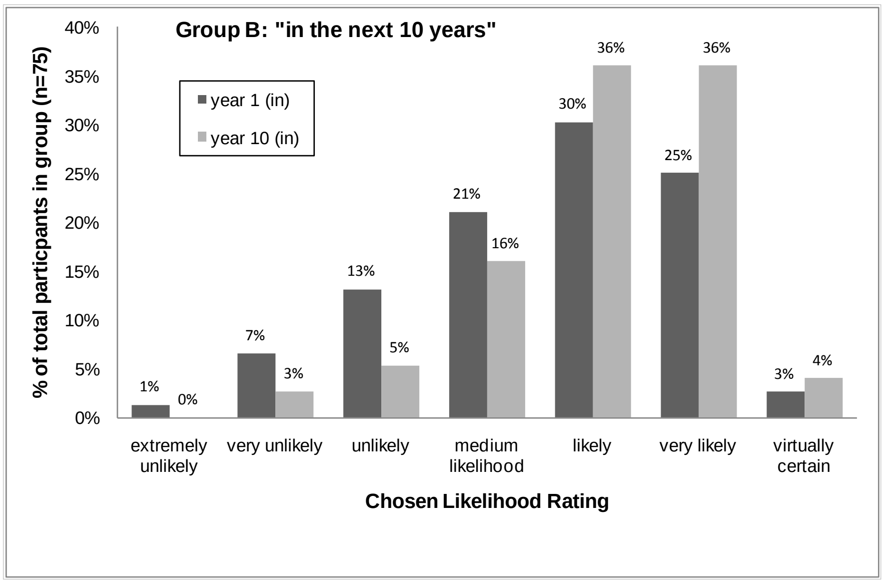
Preliminary results: A time window statement

The aim of part 2 of the survey was to explore the perception of probabilities within time window statements, and assess whether people accurately interpret the probability of an event occurring today versus a future date. For example, for the Ruaumoko statements discussed above, how do participants rate the likelihood of an eruption *today* versus in 3 days time?

In total, 7 statements were investigated, using both a within- and between- subject design. Within each experiment group questions had different likelihood ratings, probability values, and time window durations. For the first 4 randomised statements, experiment Group A received statements referring to "*within*" (followed by the number of days or years) and Group B received "*in*". For the other 3 randomized statements, one sentence feature this same assignment of "*within*" and "*in*" between groups, while for the other two sentences experiment Group A received probabilities in a numerical format while Group B received verbal terms. The IPCC3 translation table (Table 1) was used for the choice of appropriate terms and values in each group.







▲**Figure 5**: A screen shot of the online survey format for part 2, which assessed participant's likelihood ratings through time windows for multi-day statements

▶**Figure 6**: % of total participants in experiment group A ("*within*" phrasing) who rated each likelihood term for year 1 and year 10, for the question outlined in Figure 5

▼**Figure 7**: % of total participants in experiment group B ("*in*" phrasing); otherwise as for Figure 6

We report now on the results from an example statement. In this, participants were presented with a volcanic scenario and asked: '*The volcanologists state that there is a 68-88% chance of an explosive eruption in/within the next 10 years. It is the 1st of January in year 1. Rate the likelihood of an explosive eruption occurring*' (where in or within was used as appropriate).

The rating scale available was a Likert type verbal likelihood scale using the terms from the IPCC3 (Table 1) and participants were asked to rate the likelihood this year (year 1), and in year, 3, 5, 8, 10 and 15.

Phrasing	Year 1		Year 3		Year 5		Year 8		Year 10		Year 15	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Group A: within	4.71	1.073	4.79	1.006	4.85	0.959	4.9	1.077	4.89	1.193	4.55	1.436
Group B: in	4.6	1.315	4.76	1.149	4.92	1.05	5.12	1.033	5.12	1.046	4.77	1.342

▲**Table 4**: The Mean Likelihoods Ratings with Standard Deviations for Each Year, as Rated by Group A and Group B of the Time Window Investigation. We assume likelihood scale ranges from extremely unlikely=1 to virtually certain=7.

Figure 6 illustrates the rating of likelihoods for Year 1 and Year 10, for all participants in Group A that received the term "*within*" in regard to the time window.

Figure 7 illustrates the rating likelihood for Group B that received the term "*in*", with the mean likelihood ratings for both groups and all years given in Table 4. For Group A (*within*), year 10 was ranked as being of higher likelihood than year 1 by 16 participants, as lower by 8, and tied for the remaining 48 complete ratings. For this group, a Wilcoxon Signed Ranks Test did not show any significant difference between the likelihood ratings in year 1 and year 10 (Z=-1.333, *p*=0.183). In comparison, for Group B ("*in*"), a Wilcoxon Signed Ranks Test showed a significant difference between the likelihood ratings in year 1 and year 10 (Z=-3.250, *p*=0.01), with year 10 rated as higher by 23 participants (vs. 6 cases as lower, and 45 cases tied).

This is of particular interest because the numerical rules of probability indicate that the likelihood is equal in both year 1 and year 10. This preliminary result appears to indicate that many participants are not viewing the likelihood of an eruption as being uniform throughout the time window, but rather view the likelihood today as being lower. In addition, the subtle change of using "*within*" instead of "*in*" results in a more uniform distribution of the likelihood ratings through the time window. Indeed for the "*within*" condition, the total of the negative (year 1 < year 10) and positive (year 10 > year 1) ranks is 104 and 196, respectively. Meanwhile, for the "*in*" condition, the total is 70 and 365, further illustrating the higher likelihood ratings towards the end of the time window when this phrasing is used.

Discussion and Conclusions

The effective use of science advice in emergency management is fundamentally dependent upon good relationships between science advisers and key decision makers that includes effective processes for the delivery of this advice, the generation of trust and confidence in the advisors, and the effective communication of the advice in a manner and format that can be both understood and translated into effective action. Exercise Ruaumoko and other volcanic crises indicate that there is a need to identify the different ways emergency managers and volcanologists understand and use uncertainty and probabilities. In addition, not only may these decisions be affected by people's differing perception of the wording, but they may also be affected by both their perceptions of the likelihood distribution within a time window, and their understanding of how decision making thresholds relate to these time windows.

Our preliminary results from a survey tool delivered to assess the differing perceptions between scientists and non-scientists for the translation of verbal likelihood phrases appear to show that the IPCC3 and IPCC4 tables are not appropriate for use in volcanic crises. Poor translation performance for both groups, especially for the extreme values, supports the approach of building a translation table unique to the volcanological community, built up from the non-scientist community as this is the target community for communications. As discussed by Budescu et al. (2009), translation tables should still be used with caution, as they may not correspond with people's intuitive translations, and thus verbal and numerical terms and phrases should be communicated together in statements to mitigate this issue. This requires further investigation, firstly because we have currently only considered the translation from verbal to numerical terms and not vice versa, and secondly because volcanology is a field characterised by very low probability but high impact events, as of earthquakes. It is also worth noting that a contributing factor to the poor translation performance at extremes may have been due to the sliding bar scales used in this study, and that of Budescu et al. (2009).

The 'shifting' of the likelihood ratings to the end of the time window in the preliminary results from the example statement, in part 2 of the survey could result in delayed action during a volcanic crisis. The subtle change in perception across the time window due to the use of the term "*within*" instead of "*in*" in regard to the time window, highlights the care that should be used in the generation of these statements and the necessity of ensuring consistency across all statements within a crisis. In addition, it also offers a potential solution to the 'shifting' of the likelihood. However, analysis of the other statements in parts 2 and 3 of the survey must be completed before full conclusions can be drawn. In addition, it has been suggested that these statements could be tested with a positive outcome (e.g., the likelihood of winning the lottery in/within the next two weeks) to identify whether this 'shifting' may be a general cognitive displacement.

There is also scope for developing new ways of translating science to practice. For example, techniques exist within the organisational strategic management literature to help people deal with uncertainty by creating a smaller set of options for response and identifying the precursors that can identify which option is most likely to occur. This approach could also help provide a context to enhance the quality of relationships between scientific and emergency management agencies.

In conclusion, the lessons learnt from volcanic crises for the communication of uncertainty and probabilities to emergency managers, key decision makers, public officials and the community, can be applied to all natural hazards, particularly those typified by high levels of uncertainty during lead up periods to an event. Bringing this back to the September 2010 Canterbury and February 2011 Christchurch earthquakes, when we consider aftershock advice such as "… *a 23 per cent probability of a magnitude-6.0 to 6.9 event somewhere in the Canterbury aftershock zone over the next 12 months* …" (GNS Science Statement issued on 3/6/11, as reported in "Little change to risk of big quake - expert," 2011) it is clear that we must utilise the lessons from the judgment literature to format these statements in such a way as to enhance people's understanding of their content and meaning.

References

Aspinall, W., & Cooke, R. (1998). Expert judgement and the Montserrat Volcano eruption. In A. Mosleh & R. A. Bari (Eds.), *Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management PSAM4, September 13th-18th* (Vol. 3, pp. 2113–2118). New York, USA.

Blake, S., Wilson, C. J. N., Smith, I. E. M., & Leonard, G. S. (2006). *Lead times and precursors of eruptions in the Auckland Volcanic Field, New Zealand: indications from historical analogues and theoretical modelling*. GNS Science Report, # 34.

Blong, R., & McKee, C. (1995). *The Rabaul eruption 1994: Destruction of a town*. Natural Hazards Research Centre, Macquarie University, Australia.

Bruine De Bruin, W., Fischhoff, B., Millstein, S., & Halpern-Felsher, B. (2000). Verbal and numerical expressions of probability: "It's a fifty-fifty chance." *Organisational Behavior and Human Decision Processes*, *81*, 115-131.

Brun, W., & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? *Organisational Behavior and Human Decision Processes*, *41*, 390-404.

Budescu, D. V., Broomell, S., & Por, H.-H. (2009). Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science*, *20*, 299-308.

Budescu, D. V., Karelitz, T. M., & Wallsten, T. S. (2003). Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making*, *16*, 159-180.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1-73.

Cronin, S. J. (2008). The Auckland Volcano Scientific Advisory Group during Exercise Ruaumoko: Observations and recommendations. Data collection. *Civil Defence Emergency Management: Exercise Ruaumoko*. Auckland: Auckland Regional Council.

Dhami, M. K., & Wallsten, T S. (2005). Interpersonal comparison of subjective probabilities: toward translating linguistic probabilities. *Memory & Cognition*, *33*, 1057-1068.

Doyle, E. E., & Johnston, D. M. (2011). Science advice for critical decision-making. In D Paton & J. Violanti (Eds.), *Working in high risk environments: Developing sustained resilience* (p. in press). Springfield, IL: Charles C Thomas.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *British Medical Journal*, *327*, 741-744.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684-704.

Gigerenzer, G., Hertwig, R., Broek, E. V. D., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30 % chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, *25*(3), 623-629.

Haynes, K., Barclay, J., & Pidgeon, N. (2008). Whose reality counts? Factors affecting the perception of volcanic risk. *Journal of Volcanology and Geothermal Research*, *172*(3-4), 259-272.

Hill, D. P. (1998). 1998 SSA meeting - Presidential Address: Science, geologic hazards, and public in a large, restless caldera. *Seismological Research Letters*, *69*(5), 400-404.

Honda, H., & Yamagishi, K. (2006). Directional verbal probabilities. *Experimental Psychology*, *53*(3), 161-170.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., et al. (Eds.) (2002). *Climate change 2001: The scientific basis*. Cambridge, UK: Cambridge University Press.

IAVCEI Subcommittee for Crisis Protocols. (1999). Professional conduct of scientists during volcanic crises. *Bulletin of Volcanology*, *60*, 323-334.

Johnston, D. M., Scott, B., Houghton, B. F., Paton, D., Dowrick, D., Villamor, P., & Savage, J. (2002). Social and economic consequences of historic caldera unrest at the Taupo volcano, New Zealand, and the management of future episodes of unrest. *Bulletin of The New Zealand Society For Earthquake Engineering*, *35*(4), 215-230.

Joslyn, S. L., & Nichols, R. M. (2009). Probability or frequency? Expressing forecast uncertainty in public weather forecasts. *Meteorological Applications*, *16*(3), 309-314. Joslyn, S. L., Nadav-

Greenberg, L., Taing, M. U., & Nichols, R. M. (2009). The effects of wording on the understanding and use of uncertainty information in a threshold forecasting decision. *Applied Cognitive Psychology*, *23*, 55-72.

Karelitz, T. M., & Budescu, D. V. (2004). You say "probable" and I say "likely": Improving interpersonal communication with verbal probability phrases. *Journal of Experimental Psychology. Applied*, *10*, 25-41.

Kuhberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organisational Behavior and Human Decision Processes*, *75*, 23-55.

Lee, B. Y. (2010). Working together, building capacity – A case study of civil defence emergency management in New Zealand. *Journal of Disaster Research*, *5*(5), 565-576.

Leonard, G., Johnston, D. M., Paton, D., Christianson, A., Becker, J., & Keys, H. (2008). Developing effective warning systems: Ongoing research at Ruapehu volcano, New Zealand. *Journal of Volcanology and Geothermal Research*, *172*(3-4), 199-215.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organisational Behavior and Human Decision Processes*, *76*(2), 149-88.

Lindsay, J., Marzocchi, W., Jolly, G., Constantinescu, R., Selva, J., & Sandri, L. (2009). Towards real-time eruption forecasting in the Auckland Volcanic Field: application of BET_EF during the New Zealand National Disaster Exercise "Ruaumoko." *Bulletin of Volcanology*, *72*(2), 185-204.

Lipkus, I. M. (2010). Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Medical Decision Making*, *27*(5), 696-713.

Little change to risk of big quake - expert. (2011, June 3). *The Press*. Retrieved from http://www.stuff.co.nz/the-press/news/christchurch-earthquake-2011/5094308/Little-change-to-risk-of-big-quake-expert

MCDEM. (2008a). *Response Management: Director's Guideline for CDEM Group and Local Controllers [DGL06/08]*. *Group*. Wellington, NZ: Author.

MCDEM. (2008b). *Exercise Ruaumoko ' 08 Final Exercise Report* (pp. 1-79). Wellington, NZ: Author.. Retrieved from http://www.civildefence.govt.nz/memwebsite.nsf/Files/National Exercise Programme/$file/ExRuaumoko-FINAL-REPORT-Aug08.pdf

MCDEM. (2009). *CDEM Exercises Director's Guideline for Civil Defence Emergency*. *Management*. Wellington, NZ: Author.

Marzocchi, W., & Woo, G. (2007). Probabilistic eruption forecasting and the call for an evacuation. *Geophysical Research Letters*, *34*(22), 1-4.

McDowell, S. (2008). *Exercise Ruaumoko: Evaluation Report Auckland Civil Defence Emergency Management Group*. Auckland, NZ: ACDEM. Retrieved from http://www.emergencyepl.com/files/2031990/uploaded/ACDEMG_EPL%20Report_%20tsunami%20response%20081109.pdf

McGuire, W. J., Solana, M. C., Kilburn, C. R. J., & Sanderson, D. (2009). Improving communication during volcanic crises on small, vulnerable islands. *Journal of Volcanology and Geothermal Research*, *183*(1-2), 63-75.

Miles, S., & Frewer, L. J. (2003). Public perception of scientific uncertainty in relation to food hazards. *Journal of Risk Research*, *6*(3), 267-283.

New Zealand Fire Services Commission. (1998). *The New Zealand Coordinated Incident Management System (CIMS): Teamwork in emergency management*. Wellington, NZ: Author..

Newhall, C. G., & Punongbayan, R. S. (Eds.). (1996). *FIRE and MUD eruptions and lahars of Mount Pinatubo, Philippines*. Seattle, WA: University of Washington Press. Retrieved from http://pubs.usgs.gov/pinatubo/index.html

Paton, D., & Auld, T. (2006). Resilience in emergency management: Managing the flood. In D. Paton & D. Johnston (Eds.), *Disaster resilience: An integrated approach*. Springfield, IL: Charles C Thomas.

Paton, D., Johnston, D. M., & Houghton, B. F. (1998). Organisational response to a volcanic eruption. *Disaster Prevention and Management*, *7*(1), 5-13.

Patt, A., & Dessai, S. (2005). Communicating uncertainty: Lessons learned and suggestions for climate change assessment. *Comptes Rendus Geosciences*, *337*(4), 425-441.

Patt, A., & Schrag, D. P. (2003). Using specific language to describe risk and probability. *Climatic Change*, *61*, 17-30.

Risbey, J. S., & Kandlikar, M. (2007). Expressions of likelihood and confidence in the IPCC uncertainty assessment process. *Climatic Change*, *85*(1-2), 19-31.

Smith, R. (2009). Research, science and emergency management: Partnering for resilience. *Tephra, Community Resilience: Research, Planning and Civil Defence Emergency Management* (pp. 71-78). Wellington, NZ: Ministry of Civil Defence & Emergency Management.

Sparks, R. S. J. (2003). Forecasting volcanic eruptions, *Earth and Planetary Science Letters*, *210*(1-2), 1-15.

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organisational Behavior and Human Decision Processes*, *80*(2), 155-190.

Tilling, R. I. (2008). The critical role of volcano monitoring in risk reduction. *Advances in Geosciences*, *14*, 3-11.

Wallsten, T. S, Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations frequency expressions of probability. *Journal of Memory and Language*, *25*, 571-587.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin*, *115*, 228-242.

Wiedemann, P., Borner, F., & Schütz, H. (2008). Lessons learned: Recommendations for communicating conflicting evidence for risk characterization. In P. M. Wiedemann & H. Schütz (Eds.), *The role of evidence in risk characterisation: Making sense of conflicting data* (pp. 205-213). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co..

Windschitl, P. D., & Weber, E. U. (1999). The interpretation of "likely" depends on the context, but "70%" is 70% -- right? The influence of associative processes on perceived certainty. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 1514-1533.

Woo, G. (2008). Probabilistic criteria for volcano evacuation decisions. *Natural Hazards*, *45*(1), 87-97.

Author Note

Dr Emma Doyle, e.e.doyle@massey.ac.nz, is supported by Foundation for Research Science & Technology NZS&T Postdoctoral Fellowship MAUX0910. All research conducted within this study was done in accordance with Massey University's Code of Ethical Conduct for Research, Teaching and Evaluations Involving Human Participants.