International Journal of Disaster Risk Reduction 76 (2022) 103021

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

Investigating the decision thresholds for impact-based warnings in South East Asia

Sarah C. Jenkins [a,b,*], Agie W. Putra [c], Sefri Ayuliana [c], Riefda Novikarany [c], Norhadizah M. Khalid [d], Che Siti Noor Che Mamat [e], Lorenzo A. Moron [f], Maria Cecilia A. Monteverde [f], Esperanza. O. Cayanan [f], Rebecca Beckett [g], Adam JL. Harris [a]

[a] Department of Experimental Psychology, University College London, 26 Bedford Way, London, WCH1 0AP, United Kingdom
[b] Department of Psychology, Royal Holloway, University of London, Wolfson Building, Egham, Surrey, TW20 0EX, United Kingdom
[c] Public Weather Services Center, Indonesian Agency for Meteorology, Climatology and Geophysics, Jakarta, 10720, Indonesia
[d] MetMalaysia, Jabatan Meteorologi Malaysia Jalan Sultan, 46200, Petaling Jaya, Selangor, Malaysia
[e] National Disaster Management Agency, Prime Minister's Department, Level 6 & 7, Block D5, Complex D, Federal Government Administrative Complex, 62502, Putrajaya, Malaysia
[f] Department of Science and Technology, Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA), PAGASA Science Garden Complex, BIR Road, Brgy, Central, Quezon City, Metro Manila, 1100, Philippines
[g] Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB, United Kingdom

A R T I C L E I N F O

Keywords:
Impact-based forecasting
Impact-based warnings
Thresholds
Natural hazards
Decision making
Risk communication

A B S T R A C T

The move towards impact-based forecasting presents a challenge for forecasters, who must combine information not just on what the weather might *be*, but also on what the weather might *do*. Such forecasts require an integration of both likelihood and impact severity information to issue a particular weather warning. The current pre-registered study focusses on forecasters' and stakeholders' thresholds for determining the level of impact-based warnings, set in an area of the world particularly susceptible to extreme weather events. Set in the context of one hazard (heavy rainfall or river flooding), forecasters and stakeholders from Indonesia, Malaysia and the Philippines provided hypothetical impact-based warnings for impacts of varying likelihood. Results indicated generally good alignment with the warnings implied by previously developed impact tables. In the one country where a comparison was possible (the Philippines), we did not find evidence to suggest that forecasters and stakeholders use different thresholds for issuing warnings. We suggest that warning thresholds should be subject to regular monitoring wherever in the world an impact-based approach is used.

1. Introduction

Within the global risks landscape, extreme weather events are perceived as the top risk in terms of likelihood, and eighth in terms of impact [1]. During 2000–2019, natural hazards such as tsunamis, volcanoes, earthquakes, cyclones, floods and landslides caused 1.23 million deaths, affected over 4.03 billion people and led to economic losses of $2.97 trillion [2]. Whilst many of these events are well forecast with timely warnings issued, the far-reaching, adverse consequences for communities still persist. One proposed approach to

* Corresponding author. Royal Holloway, University of London, Wolfson Building, Egham, Surrey, TW20 0EX, UK.
E-mail address: sarah.jenkins@rhul.ac.uk (S.C. Jenkins).

https://doi.org/10.1016/j.ijdrr.2022.103021
Received 6 July 2021; Received in revised form 4 March 2022; Accepted 29 April 2022
Available online 7 May 2022
2212-4209/

reducing the vast impact of these natural hazards is to lessen the gap between forecasts/warnings of weather events and the understanding of their potential impacts. This gap has been observed for both the general public, but also for civil protection and emergency management authorities (World Meteorological Organization [[3]. In a bid to address this gap, there has been a paradigmatic shift towards multi-hazard impact-based forecasting (IBF), whereby forecasts include information not just on what the weather might *be*, but also on what the weather might *do* [3,4]. Such impact-based forecasts and warnings are issued with the intention that by anticipating impacts, more effective pre-emptive/preparatory mitigative action can be undertaken to minimise adverse consequences associated with these hazards. Given the high stakes nature of such forecasts, it is vital that these are grounded in the best meteorological modelling. However, it is not enough merely to improve meteorological modelling. The best forecasts in the world will mean little if the forecasts and warnings are not communicated, and understood, as intended. Improving the effectiveness of weather warnings thus requires a holistic approach – combining both physical and social science approaches.

Traditionally, weather warnings are based on weather-based factors, such as wind speed of at least x mph, or at least x cm of rainfall in a specific time period, usually expressed as a probability of this specific threshold being met or exceeded. In effect, such warnings only really consider the *hazard*, whereas impact-based warnings also involve a consideration of *exposure* ("who and what may be affected in an area in which hazardous events may occur" [[3]; p. 4]) and *vulnerability* ("susceptibility of exposed elements … to suffer adverse effects when affected by a hazard" [[3]; p. 4]), which may change over time. The same 'hazard' (e.g., wind of a particular speed and duration) will be more impactful in a densely populated area (i.e., high exposure) with poorly constructed buildings that are unable to withstand high wind speeds (i.e., high vulnerability) than in a sparsely populated (low exposure) area with robust buildings constructed to withstand high winds (low vulnerability). It is therefore vital to have detailed information on exposure and vulnerability (in addition to hazard information), so as to identify likely impacts in order to issue impact-based warnings (IBWs) [5]. Then, one way of deciding the level of warning is to make reference to an Impact Risk Matrix which combines impact severity with likelihood in a traffic light scheme, with warnings ranging from 1 (Minor) to 4 (Take Action) [3,4] (see Fig. 1). A necessary requirement for impact forecasting is the availability of adequate impact data [5], which may be assessed in different ways, for instance algorithmically (e.g., Refs. [6–8], or from impact libraries – a collection of possible impacts associated with a specific hazard (see Supplementary Materials, Table S1-Table S3 for examples) and classified according to their severity level (minimal/very low, minor/low, significant/medium, severe/high) [9,10]. Whilst more complex scientific techniques to assess impacts continue to be developed, determining the severity of impacts is still typically a subjective undertaking. Gaining an understanding of how these decisions are made is crucial to ensuring the effectiveness of impact-based forecasts.

1.1. *Investigating the effectiveness of impact-based forecasts*

Research investigating the utility of IBF has typically compared IBWs to traditional threshold-based warnings, examining their influence on understanding, risk perceptions and behavioural intentions [11–19]. Studies typically use hypothetical scenarios, in which participants are placed in a role, such as that of a resident or factory worker, and asked to decide on a series of actions. A few studies have sought to use more naturalistic, field methods when comparing warnings, for instance by using a smartphone application from a weather provider [20,21]. The results of research on IBWs is largely mixed – some studies report increased levels of understanding, increased intentions to engage in protective behaviours [11–13,19] and increased risk perceptions (levels of worry and fear, [14], for IBWs. Conversely, other studies have found no effect of IBWs on understanding, risk perceptions or behavioural intentions/responses versus standard warnings [16,17,20,21]. Much of this existing research has generally measured understanding using self-report measures, such as 'ease of understanding', whereby participants are asked to rate perceived comprehensibility of the warning. Potter et al. [17]; for example, asked participants to indicate their level of agreement regarding the statement: "I find it easy to understand the possible effects of the weather event" on a five-point Likert scale ranging from 'strongly disagree' to 'strongly agree'. Although Taylor et al. [22] measured objective understanding (i.e., by measuring agreement with six statements which had an objectively correct answer), agreements were far more common than disagreements for all statements (three of which were true, three of which were false), possibly suggesting a degree of response bias.

The ambiguity in findings from existing research thus clearly illustrates the need to further develop novel, direct measures for investigating user understanding of IBWs, as employed in the current study. Moreover, there is a pressing need for research to widen the definition of 'user.' The majority of existing research has focused on the general public as forecast users, rather than considering forecasters or stakeholders (used here to refer to disaster managers, civil protection agencies, emergency services etc) who are less far along in the warning value chain process – the chain of organisations involved in producing a warning [23,24]. The benefits of IBF are frequently cited for stakeholders, tasked with making timely effective decisions to mitigate impacts for all [5,25]. The current study addresses how (one element of) IBWs are used and understood by forecasters and key forecast users (stakeholders).

1.1.1. *The importance of alignment of impact thresholds*

Effective communication requires two components – firstly, the communicator must be perceived as a credible source of information. Secondly, for information to be used effectively, the audience must understand the information as the communicator intended [26]. Identifying whether IBWs align with those implied by the original impact table is therefore instructive for ensuring forecasts are effectively communicated. Forecasters must be able to communicate their specialist knowledge to stakeholders, such that those stakeholders then have the most accurate information readily accessible with which to plan mitigation actions. This requires an alignment of threshold values across forecasters and stakeholders. As well as aiding efficient mitigation action, threshold alignment is

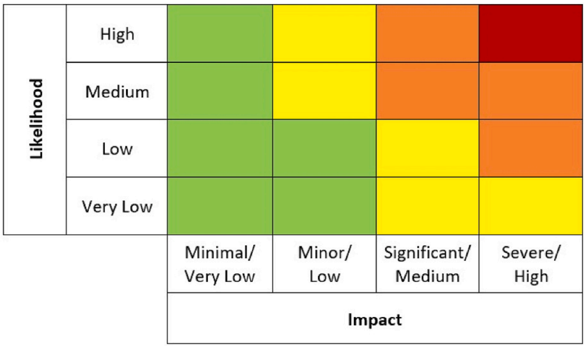

Fig. 1. Impact risk matrix.

also likely to increase stakeholders' trust in the forecasters. Meteorological forecasts rely on probabilistic predictions. Consequently, forecasters must determine the costs associated with 'misses' (where a warning is not issued and a high impact weather event occurs) and 'false alarms' (where a warning is issued, but a high impact weather event does not subsequently occur).[1] Both misses and false alarms (which may result from well-calibrated forecasters) can reduce subsequent trust in a forecaster, even if they do not happen repeatedly [27–34]. Such a loss in trust also has potential implications for future behaviour, with misses and false alarms found to be associated with reduced public responsiveness to future warnings (measured by behavioural intentions, [18,32]. Whilst both misses and false alarms *can* occur in well-calibrated forecasters whose thresholds are aligned with those of stakeholders, an alignment of the thresholds of forecasters and stakeholders provides one line of defence against their perception as a miss or false alarm, and subsequent losses of trust. This point is best made with an example: Consider a forecaster with a liberal threshold for a high impact event, forecasting the likelihood of such an event to be 'high', as per the impact matrix. An event with the impacts envisioned by the forecaster subsequently occurred. Such a forecast may nevertheless be perceived as a false alarm by a stakeholder *if* they had a more conservative threshold for a high impact weather event, and the resulting impacts did not meet their threshold. Consequently, an alignment of thresholds between forecasters and stakeholders will maximise the appropriateness of stakeholders' mitigation actions, whilst being conducive to the maintenance of stakeholders' trust in forecasters.

1.2. Current study

Despite the global rollout of the impact-based approach to forecasting, little research exists on the decision processes behind how warnings are assigned, nor on how they are understood by those other than the general public. We sought to address this gap in an area of the world particularly susceptible to extreme weather events – South East Asia [35]. In Indonesia, over the past 40 years, an average of 300 disasters occurred per year, with an average death toll of around 8000 [36]. In the last two decades, Malaysia has experienced 51 natural hazard events, with a death toll of 281 [37]. Since 1990, the Philippines has been affected by 565 natural hazard events, with a death toll of just under 70,000 [38]. However, the development and use of IBWs is still relatively new in these countries, a factor which the Weather and Climate Science for Service Partnership (WCSSP) South East Asia project aims to address by forming strong, sustainable science and innovation partnerships [39].

Focusing on forecasters and stakeholders from Indonesia, Malaysia and the Philippines, we here investigate how likelihood and impact information from impact tables is understood and used to assign IBWs, specifically relating to the alignment of decision thresholds.

The study had two main objectives:

1. Compare the (hypothetical) impact-based warnings issued by forecasters and stakeholders with those implied by the original impact table.
2. Compare the thresholds for issuing warnings between forecasters and stakeholders.

2. Method

We conducted the same study design across the three countries and thus in the following, we report the methodology generally, highlighting differences between countries as appropriate. Authors wishing to view the methods, separated by country, are directed to the pre-registration document at: https://osf.io/rkz7v/?view_only=08f20763e7a2416caafec0085390cb39 (Study 2).

[1] It is important to note that misses and false alarms are both inevitable for probabilistic predictions, and thus should not be considered errors.

Table 1
Participant information for partner countries.

Country (Partner)	Full completions (n)	Demographics		Data Collection Period
		Forecasters	Stakeholders	
Indonesia (BMKG)	96	89 44 male, 43 female, 2 prefer not to say, aged between 17 and 45 (Mdn = 27.5, one missing) IBF experience: 21.3% little or no experience; 31.4% some training; 42.6% some experience; 4.5% a lot of experience.	7–5 male, 2 female aged between 21 and 47 (Mdn = 21) Approached those from: National Disaster Management Agency, Regional Disaster Management Agency Jakarta Role: 28.5% disaster management, 71.4% 'use forecasts' IBF experience: 42.9% little or no experience; 42.9% some experience; 14.2% a lot of experience.	November 22nd 2020–January 20th 2021
Malaysia (MetMalaysia)	57 (one unspecified job role)	51 22 male, 26 female, 3 prefer not to say, aged between 26 and 49 (Mdn = 38) IBF experience: 50.9% little or no experience; 15.7% some training; 31.4% some experience; 2.0% a lot of experience.	5–1 male, 4 female, aged between 30 and 41 (Mdn = 39) Approached those from: National Disaster Management Agency Drainage and Irrigation Department Role: 60.0% disaster management, 40.0% 'use forecasts' IBF experience: 40% little or no experience; 60% some experience.	November 27th 2020–January 20th 2021
Philippines (PAGASA)	54	28 15 male, 13 female, aged between 25 and 62 (Mdn = 32.5) IBF experience: 39.2% little or no experience; 32.1% some training; 28.6% some experience.	26–16 male, 10 female, aged between 22 and 62 (Mdn = 35) Approached: DRRM Officers from several Local Government Units in Metro Manila and Metro Cebu Role: 92.3% disaster management, 3.8% emergency services, 3.8% 'use forecasts' IBF experience: 34.6% little or no experience; 42.3% some training; 23.1% some experience.	November 11th 2020–January 20th 2021

2.1. Participants

Forecasters and stakeholders were recruited by the South East Asian authors to participate in this online study. A description of the study and the survey link was emailed to those identified by the local authors, with reminder emails sent periodically throughout the data collection period (indicated in Table 1).[2] The Philippines also carried out an IBF training seminar during the data collection period, attended by stakeholders from several Local Government Units (LGUs) in Metro Manila and Metro Cebu, where they were also approached to participate. For full demographic details of each sample, see Table 1. Participation in the study was not remunerated. Ethical approval was granted from the Departmental Ethics Chair for Experimental Psychology (University College London).

2.2. Design

The design of the study was based on the World Meteorological Organisation's [3] Impact Risk Matrix for impact-based weather warnings (see Fig. 1). Impact severity (four levels – minimal, minor, significant and severe[3]) was manipulated within-participants by picking impacts from the relevant sections of each country's impact tables (see Supplementary Materials, Table S1-Table S3). Our design focussed on forecasts made when likelihood is 'high', given that it is only when likelihood is 'high' that four different warning levels are prescribed for each of the four different impact categories (see Fig. 1). Consequently, all impacts were presented as having high likelihood at some point in the task. Given that impact tables already existed within the three countries, we were interested in the recovered impact severity thresholds when participants combined likelihood and impact severity information. To reinforce this task requirement and also increase ecological validity, a random selection of impacts were presented twice (see Table 2), but the second time presented as having either a medium, low or very low likelihood. This random selection of impacts was interspersed within the task. Participant group (forecaster or stakeholder) was a between-participants variable.

2.3. Materials

Impacts were presented in the context of a specific hazard, selected in collaboration with the South East Asian authors on the basis of its high relevance to that country and the existence of a pre-developed impact table (see Table 2). Each impact and its likelihood level was individually presented, one at a time in a randomised order. Participants were asked to assign a warning level, on a scale of 1 (Minor) to 4 (Take Action; see Fig. 2). On the advice of the local authors, materials were presented in English in Malaysia and the Philippines, and translated into Bahasa in Indonesia. Although a complete back translation was not possible, the translation was undertaken by SA and RN and checked by AW, with random selections additionally checked by RB and SJ.

2.4. Procedure

The study was run online via Qualtrics, with participants able to complete the study in more than one session. Before beginning the tasks, participants were asked a series of demographic questions. They were asked to indicate: gender (male/female/prefer not to say); age; their area of work (meteorology/disaster management/civil protection/emergency services/other – Please indicate if, in your work, you typically use forecasts/or prepare them) and their level of experience with IBF (little or no experience/some training/some experience/a lot of experience).

Participants first completed a related study using the psychometric paradigm (see https://osf.io/rkz7v/'Study 1 – Impact Perceptions'[4]). In this, participants were asked to rate severe impacts on a number of characteristics, including their destructiveness, scope of area affected and predictability, in relation to a specific hazard (specified in Table 2).

Participants were then presented with instructions for the second (current) task, asking them to consider the presented impact and likelihood information, and indicate which warning level they would assign. They were asked to consider the impacts solely in relation to their country.

On the next screen, participants were presented with one of the impacts and a specified likelihood level and asked to assign a warning, from one of the four labels (see Fig. 2), with the impact risk matrix provided below for context. The impacts were presented in a random order to avoid order effects. The subsequent screen showed another randomly presented impact and likelihood level, and so on and so forth, until the participant had rated the full set of impacts. Finally, participants were thanked and debriefed.

3. Results

We specified a minimum number of participants (n = 30) for each group (forecasters and stakeholders in each country) in our pre-registration, but this was not reached in the Indonesian or Malaysian samples (seven and five stakeholders, respectively). We therefore only conducted analysis on the forecaster group in these samples.[5] Although we fell slightly short of this number in the Filipino sample (28 forecasters, 26 stakeholders), this was far more than the number of stakeholders recruited in either of the other two countries and

[2] We pre-registered that we would aim to collect as many participants as possible (minimum of 30 per group) within a specified timeframe (four weeks). The specialist nature of the sample meant there were limited numbers of individuals to recruit from at each organisation. In light of this, and following feasibility discussions with the South East Asia authors, we settled on 30 participants. Given our difficulties with obtaining this minimum number of participants after four weeks, we subsequently extended the data collection period. No data analyses were undertaken prior to the full reported samples being recruited.

[3] In the Philippines, impact severity was presented with the following four levels – minimal/very low, minor/low, significant/medium, severe/high – See Fig. 1.

[4] This study was written up separately, owing to its different aims and methodology.

[5] Although we did not pre-register any exclusion criteria, one participant did not complete any of the questions in the Malaysian sample and was removed from analysis, leaving a total of 50 forecasters.

Table 2
Design and materials by country.

Country	Hazard	Number of Impacts	Number of Impacts Presented Twice	Total Number of Ratings
Indonesia	Heavy rainfall	43	8	51
Malaysia	River flooding	56	12	68
Philippines	Heavy rainfall	58	12	70

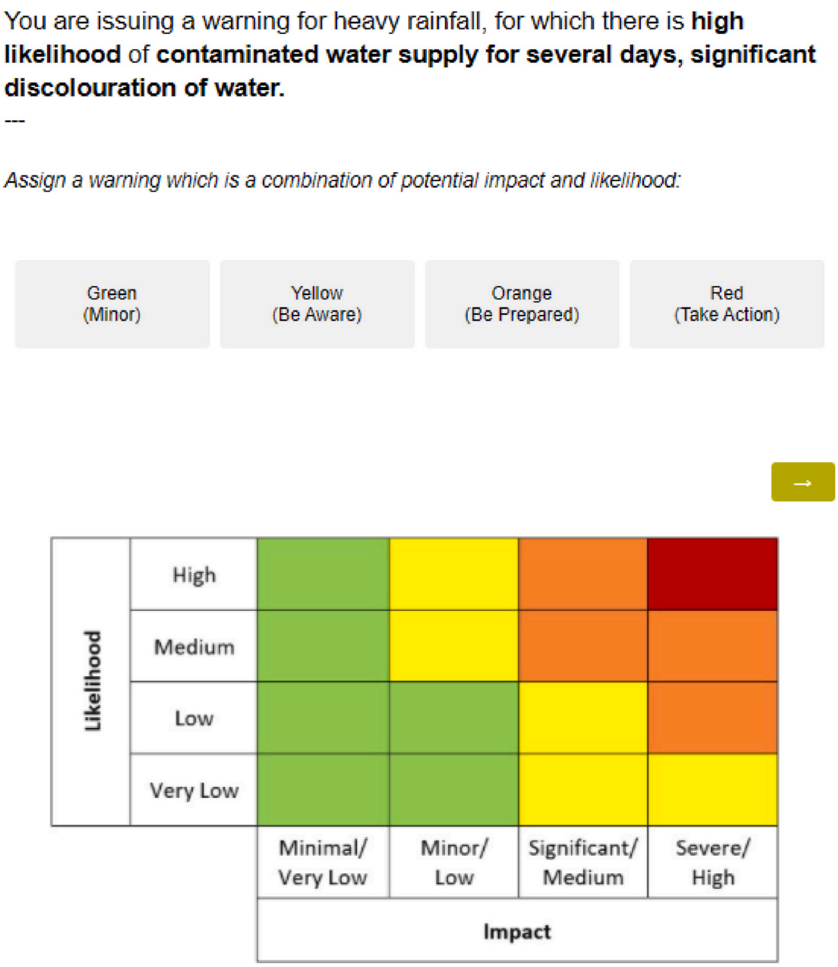

Fig. 2. Example question for Filipino participants.

roughly evenly distributed. We thus continued to conduct analyses on both groups in this sample.[6]

The following analysis only features ratings of impacts where they were presented as having high likelihood. Because we were interested in understanding how forecasters and stakeholders view every impact, so as to gain a precise understanding of any deviations from extant impact tables, or between forecasters and stakeholders, our primary analyses are at the level of individual impacts.

We coded responses such that Green (Minor) warnings were assigned a value of 1, Yellow (Be Aware) warnings assigned a value of 2, Orange (Be Prepared) warnings assigned a value of 3 and Red (Take Action) assigned a value of 4.

3.1. Objective 1. compare the (hypothetical) impact-based warnings issued by forecasters and stakeholders with those implied by the original impact table – planned analysis

We created 'alignment scores' to capture how extreme the warning was in relation to the original impact table (see Supplementary Materials, Table S1-Table S3). This was calculated as the participant's answer minus the 'original' impact categorisation, see Table 3. Following Barnes et al. [27]; we use the term 'over-warning' to refer to warnings which were of higher severity than prescribed by the

[6] No analyses were undertaken before the decision was made to stop data collection.

Table 3
Calculation of alignment score.

		'Original' Impact Categorisation			
		1	2	3	4
Participant Answer	1	0	−1	−2	−3
	2	1	0	−1	−2
	3	2	1	0	−1
	4	3	2	1	0

Note: Score of 0 represents complete alignment, a negative score represents an under-warning, and a positive score represents an over-warning.

original impact table, with the term 'under-warning' relating to warnings which were of lower severity.

We created an overall alignment score, which was an average of all of the alignment scores for each impact.[7] We subsequently conducted four, One Sample Wilcoxon Rank Sum Tests (two tailed) on overall alignment score (see Table 4). Generally, performance in Malaysia and the Philippines was similar in so far as the three samples typically gave more over-warnings. In contrast, in Indonesia, overall alignment scores suggested forecasters generally gave more under-warnings.

3.1.1. Exploratory analysis

The overall alignment scores do not, however, give us the full picture when it comes to the impact categories, so we also investigated alignment scores for each level of impact severity. These were calculated as an average of the alignment scores for each impact within the minimal, minor, significant and severe categories. We then conducted One Sample Wilcoxon Rank Sum Tests (two tailed) on each of the four alignment scores (see Table 5).

All four samples across the three countries gave more over-warnings for minimal impacts, though the extent to which this occurred differed by country. The Malaysian sample gave far larger over-warnings (87.5% of impacts had a median score of 3, and the rest had a median score of 2) compared to the Indonesian and Filipino samples. In Indonesia, 100% of impacts had a median of 2, and in the Filipino sample, the majority of impacts had a median score of 2 or 3.

Similarly, all four samples gave more over-warnings for minor impacts, though to a lesser degree than for minimal impacts. The extent of over-warnings was smallest in the Indonesian sample, followed by the Filipino samples, with the largest over-warnings occurring in the Malaysian sample. Only a minority of impacts (8%) in the Malaysian sample had a median score of 2.

The samples differed slightly more by country in their warnings for significant impacts. The Filipino sample showed good alignment with the impact table. The Indonesian sample gave more under-warnings with just over half of impacts with a median of 3 and the rest a median of 2. Again, the Malaysian sample gave more over-warnings, issuing higher warnings than implied by the impact table.

Finally, all of the samples gave under-warnings for severe impacts, though again the extent of this differed by country. The Indonesian sample seemed reluctant to issue the highest warning at all – only 12% of impacts had a median of 4. In contrast, the Malaysian and Filipino samples under-warned to a lesser degree – the majority of impacts in the Malaysian and Filipino forecasters had a median of 4 (83.3% and 100%, respectively), though this was 50% for the Filipino stakeholders.

Overall, with the exception of Malaysian participants' warnings for severe impacts, the pattern of any areas is in the expected direction if participants are generally consistent with the impact tables. That is, *overall* one expects minimal impacts to be over-warned (and severe impacts to be under-warned) given that there is only one possible direction of error. Consequently, this regressive pattern of results (slight over-warnings for minimal and minor impacts, slight under-warnings for significant and severe impacts) is to be expected and should not be over-interpreted at the group level. It is, however, important for countries to investigate specific impacts which might be showing greater discrepancies from the impact tables (see Supplementary Materials, Table S4-Table S6). For instance, in Indonesia, the median warning for 'closure of low-lying bridges' was yellow (Supplementary Materials Table S4, when in fact the original impact tables categorise this as severe and imply a red warning. We return to the possible consequences of such discrepancies in the discussion.

3.2. Objective 2. compare the thresholds for issuing warnings between forecasters and stakeholders – planned analyses

Owing to difficulties recruiting stakeholders in Malaysia and Indonesia, we could only investigate this in the Filipino sample. We carried out a two-sided Wilcoxon Rank Sum Test on warning ratings for each of the 58 impacts, see Supplementary Materials Table S6. We only observed significant group differences ($p < .05$) for seven impacts (one in the minimal category, six in the severe category). Generally, forecasters gave more precautionary (over-) warnings than stakeholders for severe impacts, but for the minor impact, the opposite occurred. Given the relatively small sample size and the fact that these differences were no longer significant with a Bonferroni correction applied, we see the results as consistent with the notion that forecasters and stakeholders use similar thresholds for issuing warnings.

4. Discussion

Whilst existing research has investigated the utility of IBF, this has largely focused on the general public as users of the forecasts and

[7] For impacts which appeared in more than one category (i.e., 3 and 4), the 'alignment score' was calculated as zero, if the participant answered either 3 or 4 (and thus −1 if they answered 2 and -2 if they answered 1).

Table 4
Median overall alignment score by country.

Country		Mdn	*r* (effect size)
Indonesia		- 0.30***	- 0.72
Malaysia		0.48***	0.61
Philippines	Forecasters	0.24***	0.70
	Stakeholder	0.23*	0.43

Z value = −6.67 (Indonesia), 4.29 (Malaysia), 3.70 (Filipino forecasters), 2.20 (Filipino stakeholders).
***$p < .001$, **$p < .01$, *$p < .05$.

Table 5
Median alignment scores for each level of impact severity by country.

		Median Alignment Score			
		Minimal	Minor	Significant	Severe
Indonesia		1.00***	0.25***	- 0.29***	- 0.82***
Malaysia		1.88***	1.08***	0.35***	- 0.56***
Philippines	Forecasters	1.17***	0.47***	- 0.03	- 0.22***
	Stakeholder	1.28***	0.60***	0.03	- 0.38***

***$p < .001$ (from Wilcoxon Rank Sum Tests).

assessed understanding using self-report measures. In the current study, we took an entirely novel approach, focusing on forecasters and stakeholders, assessing understanding via a quantitative warning decision task. Specifically, we compared IBWs issued by forecasters and stakeholders with those implied by the extant impact tables developed in those countries. To do so, we calculated alignment scores which showed that, generally, participants in Malaysia and Philippines gave more over-warnings (i.e., issued higher warnings than implied by the impact table), whereas participants in Indonesia gave more under-warnings (i.e., issued lower warnings than implied by the impact table).

The overall alignment scores did not, however, give us the full picture when it came to specific impact categories, so we also investigated alignment scores for each level of impact severity. Generally, if participants' thresholds are approximately aligned with the impact table thresholds, but include some noise, we would expect to observe positive alignment scores (over-warnings) for minimal and minor impacts, and negative alignment scores (under-warnings) for significant and severe impacts. On the whole, this is mostly what we found, though the extent to which this occurred varied by country. Over-warnings for minimal and minor impacts were larger in Malaysia than in Indonesia and Philippines, where similar levels of over-warning were observed. It was a slightly different picture for significant impacts – participants in the Philippines were very well aligned, but whilst participants in Indonesia provided under-warnings, participants in Malaysia showed the opposite pattern. Finally, for severe impacts, warnings were associated with particularly low severity levels in Indonesia, with very few uses of the 'Take Action' warning. The extent of under-warnings was smaller in the Filipino and Malaysian samples.

Table 5 shows that minimal impacts were associated with greater over-warnings than severe impacts were associated with under-warnings. Such a pattern is consistent with an asymmetric loss function account [40]. This account proposes that when estimating uncertain events, individuals are sensitive to the consequences of mis-estimating, which are asymmetric for over-versus underestimates. The greater the disutility of the event, the more costly the errors of underestimating often are. In the context of impact-based forecasting, under-warnings could lead to a failure to take a threat seriously enough and thus failures to engage in appropriate mitigative actions with potentially fatal consequences. Such an account has been proposed as an explanation for why individuals overestimate probabilities associated with severe negative events [41]. Although overestimations may give rise to a few costs, such as unnecessary monitoring, these will often be less than those associated with underestimates. Consequently, over-warnings are a way of protecting oneself against the negative effects of underestimation and could underlie the decision thresholds used for warnings in the current experiment. Similarly, when making evacuation decisions, the belief that it is 'better to be safe than sorry' often leads people to err on the side of caution [30,42].

Close attention should be paid to the observation that Indonesian forecasters appeared reluctant to use the highest warning category (see Supplementary Materials Table S4), which could prove costly in more ways than just opportunities for taking mitigative action (as highlighted above). For instance, an event which occurs with more severe impacts than predicted could lead to reduced communicator credibility [30] and also reduced responsiveness to future warnings [18,32].

Overall, in general, both forecasters and stakeholders seemed relatively well-aligned with the impact tables in this hypothetical warning issuing task. On the basis of these results, we suggest forecasters and stakeholders are understanding IBWs approximately as intended (at least as implied by the impact tables). These encouraging results are in line with some of the previous research which has demonstrated better understanding amongst the public for IBWs versus traditional, threshold-based weather warnings [17,19]. Thus, the early signs are promising for IBWs fulfilling their intended purpose of improving the utility of weather forecasting.

4.1. Comparing forecasters and stakeholders

Reassuringly, we found little evidence of differences between forecasters and stakeholders in their thresholds for issuing warnings. This increases the likelihood that suitable mitigation action will be taken on the basis of the forecast. Of course, these are tentative conclusions, made on the results from a relatively small sample of forecasters and stakeholders, and relatively early in the development of IBF in the Philippines. It is therefore worth monitoring this in the future, to ensure alignment between the two groups is maintained. However, the current data suggest that the efforts the Philippines have made to build relationships with stakeholders including civil protection and disaster managers (see Ref. [39] seem to be paying off in terms of building a consistent situational understanding. Future strong working relationships can only be beneficial in this regard. Indeed, the WMO [3] guidelines specifically highlight that strong partnerships between meteorological organisations and relevant agencies/user communities are key to providing effective impact forecast and warning services (a point specifically emphasized in the [4] guidelines). Such collaboration will result in a range of benefits, including improved risk assessments and monitoring, early warning, enhanced responses to natural hazards, and ultimately, lives saved.

It is important to note that there is no rationale to generalise the aforementioned finding in the Philippines to Indonesia, Malaysia or, indeed, any other country adopting IBF. The efforts that the Philippines have made to build relationships between stakeholders and forecasters likely contributes to both the general alignment we observe, but also our ability to recruit stakeholders in the Philippines to enable such an analysis in the first instance – in short, a selection bias.

4.2. Further considerations

The applied nature of this research meant that we needed to recruit participants such as forecasters and stakeholders, who were working full-time and already had considerable demands on their time, which was exacerbated by the COVID-19 pandemic. We were therefore unable to recruit as many participants as we would have liked, particularly stakeholders. We also recognise that there were differences in the age of participants and the level of experience with/training in IBF within our samples. It is possible that these factors could influence results and thus follow up work should seek to recruit a larger, more diverse sample, whereby one could compare the effect of training to assess its relevance and effectiveness. Furthermore, the recruitment of non-remunerated experts also necessitated as much brevity as possible in the experimental materials. Consequently, we were only able to investigate impact perceptions for one hazard within each country, and provided very limited detail about the meteorological information on which forecasters were meant to base their forecasts. Whilst the current signs are promising, this research should be extended to investigate a range of hazards, using a more meteorologically realistic task.

The current paradigm involved the assignment of a warning on the basis of information about one impact at one specific timepoint. However, in a real-world context, a hazard will have multiple impacts, which may vary in severity and which may change over time. Future research should therefore seek to investigate how forecasters aggregate across a number of impacts to produce a single warning. For instance, investigating which warning is most appropriate given a number of cross category impacts (e.g., mostly significant but with a couple of severe impacts) – does the presence of just one or two severe impacts increase the warning level? There is also the question of how warnings are interpreted when they change over time, for instance going from a yellow to an orange warning. Whilst the current research represents a starting point to the investigation of impact-based warnings, simulating real-world contexts as closely as possible will further increase the generalisability of findings. Additionally, given resource demands, future research might consider focusing on impacts relating to human casualties (as suggested by an anonymous reviewer), in light of the higher costs of mis-interpretations associated with such impacts.

4.3. Conclusions and implications

Overall, we found warnings issued by stakeholders and forecasters were relatively well-aligned with those implied by the impact table. In addition, we did not find evidence to suggest that forecasters and stakeholders use different thresholds for issuing warnings within the Philippines. Further developing relationships between meteorological organisations and stakeholders is key to ensuring the success of impact-based forecasting. Continued monitoring of threshold perceptions in Indonesia, Malaysia and the Philippines will be beneficial to ensure communication effectiveness in this region. More generally, we argue that such monitoring should be a critical aspect of continued evaluations of impact-based forecasting practices, wherever in the world these are utilised.

Availability of data

Data can be accessed at https://osf.io/rkz7v/?view_only=08f20763e7a2416caafec0085390cb39 (Study 2 – Impact Warning Decision Thresholds).

Author contributions

AJLH conceived the study; AJLH and SCJ designed the study with input from all authors; SCJ programmed the study; AW, SA and RMM were responsible for data collection in Indonesia; NBMK and CSNBCM were responsible for data collection in Malaysia; LM, MCM and EC were responsible for data collection in the Philippines; SCJ analysed the data; AJLH advised on analysis decisions; SCJ wrote the manuscript; AJLH edited the manuscript; All authors commented on the manuscript; RB co-ordinated the international collaboration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work and its contributors (Sarah Jenkins, Adam Harris [UCL]; Agie W Putra, Sefri Ayuliana, Riefda Novikarany [BMKG]; Norhadizah Bt Mohd Khalid [MetMalaysia] and Che Siti Noor Che Mamat [NADMA]; Lorenzo A Moron, Maria Cecilia A Monteverde, Esperanza Cayanan [PAGASA] and Rebecca Beckett [UK Met Office]) were supported by the Met Office Weather and Climate Science for Service Partnership (WCSSP) South East Asia as part of the Newton Fund. Our thanks go to the in-country partners and all of our participants – without this help, the work could not have happened. We are also grateful to Hazel Napier, Tim Aldridge, Steven Cole and Joanne Robbins for their input in discussions regarding the development of this research and comments on a previous draft.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijdrr.2022.103021.

References

[1] World Economic Forum, The Global Risks Report 2021, sixteenth ed., 2021. http://wef.ch/risks2021.
[2] United Nations Office for Disaster Risk Reduction, The Human Cost of Disasters: an Overview of the Last 20 Years (2000-2019), 2020, https://doi.org/10.18356/79b92774-en.
[3] World Meteorological Organization, WMO Guidelines on Multi-Hazard Impact-Based Forecast and Warning Services, 2015.
[4] World Meteorological Organization, WMO Guidelines on Multi-Hazard Impact-Based Forecast and Warning Services - Part II: Putting Multi-Hazard IBFWS into Practice, 2021. https://reliefweb.int/report/world/wmo-guidelines-multi-hazard-impact-based-forecast-and-warning-services-part-ii-putting.
[5] B. Merz, C. Kuhlicke, M. Kunz, M. Pittore, A. Babeyko, D.N. Bresch, D.I.V. Domeisen, F. Feser, I. Koszalka, H. Kreibich, F. Pantillon, S. Parolai, J.G. Pinto, H. J. Punge, E. Rivalta, K. Schröter, K. Strehlow, R. Weisse, A. Wurpts, Impact forecasting to support emergency management of natural hazards, Rev. Geophys. 58 (4) (2020), https://doi.org/10.1029/2020RG000704.
[6] G. Aznar-Siguan, D.N. Bresch, CLIMADA v1: a global weather and climate risk assessment platform, Geosci. Model Dev. (GMD) 12 (7) (2019) 3085–3097, https://doi.org/10.5194/GMD-12-3085-2019.
[7] R. Hemingway, J. Robbins, Developing a hazard-impact model to support impact-based forecasts and warnings: the Vehicle OverTurning (VOT) Model, Meteorol. Appl. 27 (1) (2020) 1–17, https://doi.org/10.1002/met.1819.
[8] T. Röösli, C. Appenzeller, D.N. Bresch, Towards operational impact forecasting of building damage from winter windstorms in Switzerland, Meteorol. Appl. 28 (6) (2021), https://doi.org/10.1002/met.2035.
[9] T. Aldridge, O. Gunawan, R.J. Moore, S.J. Cole, G. Boyce, R. Cowling, Developing an impact library for forecasting surface water flood risk, J. Flood Risk Manag. 13 (3) (2020) e12641, https://doi.org/10.1111/jfr3.12641.
[10] Met Office, What Are the National Severe Weather Warning Service Impact Tables?, 2017. https://www.metoffice.gov.uk/weather/guides/severe-weather-advice.
[11] M.A. Casteel, Communicating increased risk: an empirical investigation of the National Weather Service's impact-based warnings, Weather, Clim. Soc. 8 (3) (2016) 219–232, https://doi.org/10.1175/WCAS-D-15-0044.1.
[12] M.A. Casteel, An empirical assessment of impact based tornado warnings on shelter in place decisions, Int. J. Disaster Risk Reduc. 30 (February) (2018) 25–33, https://doi.org/10.1016/j.ijdrr.2018.01.036.
[13] E.R. Meléndez-Landaverde, M. Werner, J. Verkade, Exploring protective decision-making in the context of impact-based flood warnings, J. Flood Risk Manag. 13 (1) (2020), https://doi.org/10.1111/jfr3.12587.
[14] R.E. Morss, C.L. Cuite, J.L. Demuth, W.K. Hallman, R.L. Shwom, Is storm surge scary? The influence of hazard, impact, and fear-based messages and individual differences on responses to hurricane risks in the USA, Int. J. Disaster Risk Reduc. 30 (September 2017) (2018) 44–58, https://doi.org/10.1016/j.ijdrr.2018.01.023.
[15] D. Mu, T.R. Kaplan, R. Dankers, Decision making with risk-based weather warnings, Int. J. Disaster Risk Reduc. 30 (March) (2018) 59–73, https://doi.org/10.1016/j.ijdrr.2018.03.030.
[16] M.F. Perreault, J.B. Houston, L. Wilkins, Does scary matter?: testing the effectiveness of new National Weather Service tornado warning messages, Commun. Stud. 65 (5) (2014) 484–499, https://doi.org/10.1080/10510974.2014.956942.
[17] S.H. Potter, P.V. Kreft, P. Milojev, C. Noble, B. Montz, A. Dhellemmes, R.J. Woods, S. Gauden-Ing, The influence of impact-based severe weather warnings on risk perceptions and intended protective actions, Int. J. Disaster Risk Reduc. 30 (March) (2018) 34–43, https://doi.org/10.1016/j.ijdrr.2018.03.031.
[18] J.T. Ripberger, C.L. Silva, H.C. Jenkins-Smith, M. James, The influence of consequence-based messages on public responses to tornado warnings, Bull. Am. Meteorol. Soc. 96 (4) (2015) 577–590, https://doi.org/10.1175/BAMS-D-13-00213.1.
[19] P. Weyrich, A. Scolobig, D.N. Bresch, A. Patt, Effects of impact-based warnings and behavioral recommendations for extreme weather events, Weather, Clim. Soc. 10 (4) (2018) 781–796, https://doi.org/10.1175/WCAS-D-18-0038.1.
[20] P. Weyrich, A. Scolobig, F. Walther, A. Patt, Do intentions indicate actual behaviour? A comparison between scenario-based experiments and real-time observations of warning response, J. Contingencies Crisis Manag. 28 (3) (2020) 240–250, https://doi.org/10.1111/1468-5973.12318.
[21] P. Weyrich, A. Scolobig, F. Walther, A. Patt, Responses to severe weather warnings and affective decision-making, Nat. Hazards Earth Syst. Sci. 20 (10) (2020) 2811–2821, https://doi.org/10.5194/nhess-20-2811-2020.
[22] A.L. Taylor, A. Kause, B. Summers, M. Harrowsmith, Preparing for Doris: exploring public responses to impact-based weather warnings in the United Kingdom, Weather, Clim. Soc. 11 (4) (2019) 713–729, https://doi.org/10.1175/WCAS-D-18-0132.1.
[23] B. Golding, M. Mittermaier, C. Ross, B. Ebert, S. Panchuk, A. Scolobig, D. Johnston, A Value Chain Approach to Optimising Early Warning Systems, 2019. https://www.undrr.org/publication/value-chain-approach-optimising-early-warning-systems.
[24] Q. Zhang, L. Li, B. Ebert, B. Golding, D. Johnston, B. Mills, S. Panchuk, S. Potter, M. Riemer, J. Sun, A. Taylor, S. Jones, P. Ruti, J. Keller, Increasing the value of weather-related warnings, Sci. Bull. 64 (10) (2019) 647–649, https://doi.org/10.1016/j.scib.2019.04.003.
[25] R. Campbell, D. Beardsley, T. Sezin, Impact-based Forecasting and Warning: Weather Ready Nations, 2018. https://public.wmo.int/en/resources/bulletin/impact-based-forecasting-and-warning-weather-ready-nations.
[26] S.C. Jenkins, Communicating Uncertainty: the Role of Communication Format in Maximising Understanding and Maintaining Credibility, University College London, 2020.
[27] L.R. Barnes, E.C. Gruntfest, M.H. Hayden, D.M. Schultz, C. Benight, False alarms and close calls: a conceptual model of warning accuracy, Weather Forecast. 22 (5) (2007) 1140–1147, https://doi.org/10.1175/WAF1031.1.
[28] R.N. Collins, D.R. Mandel, Cultivating credibility with probability words and numbers, Judgment Decis. Mak. 14 (6) (2019) 683–695, https://doi.org/10.31234/osf.io/gkduw.
[29] N.F. Dieckmann, R. Mauro, P. Slovic, The effects of presenting imprecise probabilities in intelligence forecasts, Risk Anal. 30 (6) (2010) 987–1001, https://doi.org/10.1111/j.1539-6924.2010.01384.x.

[30] S.C. Jenkins, A.J.L. Harris, R.M. Lark, When unlikely outcomes occur: the role of communication format in maintaining communicator credibility, J. Risk Res. 22 (5) (2019) 537–554, https://doi.org/10.1080/13669877.2018.1440415.
[31] S.C. Jenkins, A.J.L. Harris, Maintaining credibility when communicating uncertainty: the role of directionality, Think. Reas. 27 (1) (2020) 97–123, https://doi.org/10.1080/13546783.2020.1723694.
[32] J.T. Ripberger, C.L. Silva, H.C. Jenkins-Smith, D.E. Carlson, M. James, K.G. Herron, B. Blvd, False alarms and missed events: the impact and origins of perceived inaccuracy in tornado warning systems, Risk Anal. 35 (1) (2015), https://doi.org/10.1111/risa.12262.
[33] U. Sharma, A. Patt, Disaster warning response: the effects of different types of personal experience, Nat. Hazards 60 (2) (2012) 409–423, https://doi.org/10.1007/s11069-011-0023-2.
[34] G. Woo, Probabilistic criteria for volcano evacuation decisions, Nat. Hazards 45 (1) (2008) 87–97, https://doi.org/10.1007/s11069-007-9171-9.
[35] World Economic Forum, Why Asia-Pacific Is Especially Prone to Natural Disasters, 2018. https://www.weforum.org/agenda/2018/12/why-asia-pacific-is-especially-prone-to-natural-disasters/.
[36] Global Facility for Disaster Reduction and Recovery, Country Profile: Indonesia, 2015. https://www.gfdrr.org/en/publication/country-profile-indonesia.
[37] Centre for Excellence in Disaster Management and Humanitarian Assistance, Malaysia Disaster Management Reference Handbook, 2019.
[38] Global Facility for Disaster Reduction and Recovery, Country Profile: the Philippines, 2016. https://www.gfdrr.org/sites/default/files/publication/PHILIPPINES2016.pdf.
[39] R. Beckett, A. Hartley, Progress on the Development of Impact Based Forecasting in South East Asia, Met Office, 2020.
[40] E.U. Weber, From subjective probabilities to decision weights: the effect of asymmetric loss functions on the evaluation of uncertain outcomes and events, Psychol. Bull. 115 (2) (1994) 228–242, https://doi.org/10.1037/0033-2909.115.2.228.
[41] A.J.L. Harris, A. Corner, U. Hahn, Estimating the probability of negative events, Cognition 110 (1) (2009) 51–64, https://doi.org/10.1016/j.cognition.2008.10.006.
[42] E.E.H. Doyle, J. McClure, D. Paton, D.M. Johnston, Uncertainty and decision making: volcanic crisis scenarios, Int. J. Disaster Risk Reduc. 10 (PA) (2014) 75–101, https://doi.org/10.1016/j.ijdrr.2014.07.006.