

This is a repository copy of *Knowing what to do substantially improves the effectiveness of flood early warning*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/178221/>

Version: Published Version

Article:

Kreibich, Heidi, Hudson, Paul orcid.org/0000-0001-7877-7854 and Merz, Bruno (2021) Knowing what to do substantially improves the effectiveness of flood early warning. *Bulletin of the American Meteorological Society*. pp. 1-38. ISSN 0003-0007

<https://doi.org/10.1175/bams-d-20-0262.1>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Knowing What to Do Substantially Improves the Effectiveness of Flood Early Warning

Heidi Kreibich, Paul Hudson, and Bruno Merz

ABSTRACT: Flood warning systems are longstanding success stories with respect to protecting human life, but monetary losses continue to grow. Knowledge on the effectiveness of flood early warning in reducing monetary losses is scarce, especially at the individual level. To gain more knowledge in this area, we analyze a dataset that is unique with respect to detailed information on warning reception and monetary losses at the property level and with respect to amount of data available. The dataset contains 4,468 loss cases from six flood events in Germany. These floods occurred between 2002 and 2013. The data from each event were collected by computer-aided telephone interviews in four surveys following a repeated cross-sectional design. We quantitatively reveal that flood early warning is only effective in reducing monetary losses when people know what to do when they receive the warning. We also show that particularly long-term preparedness is associated with people knowing what to do when they receive a warning. Thus, risk communication, training, and (financial) support for private preparedness are effective in mitigating flood losses in two ways: precautionary measures and more effective emergency responses.

KEYWORDS: Adaptation; Damage assessment; Emergency preparedness; Emergency response; Flood events

<https://doi.org/10.1175/BAMS-D-20-0262.1>

Corresponding author: Heidi Kreibich, Heidi.Kreibich@gfz-potsdam.de

In final form 8 April 2021

©2021 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

AFFILIATIONS: Kreibich—Section Hydrology, GFZ German Research Centre for Geosciences, Potsdam, Germany; Hudson—Institute for Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany; Merz—Section Hydrology, GFZ German Research Centre for Geosciences, and Institute for Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany

Flood early warning systems aim to protect human life and reduce monetary losses [United Nations Development Programme (UNDP); UNDP 2018]. Investments in the development and implementation of warning systems for natural hazards were recommended by the 2005 United Nations (UN) Hyogo Framework for Action and this advice was renewed in the 2015 Sendai Framework for Disaster Risk Reduction [United Nations Office for Disaster Risk Reduction (UNISDR); UNISDR 2015]. Flood warning systems have a long history; for instance, a flood warning system was established along the rivers of central Germany already in 1889 (Deutsch and Pörtge 2001). Flood warning systems have now become an essential part of integrated flood risk management (Stephens and Cloke 2014). A flood early warning system consists of the interacting components of risk knowledge, monitoring and forecasting, warning dissemination and communication, and response capabilities. Several organizations and stakeholders, including the public, must be able to contribute to, and act upon these components (Perera et al. 2019). For instance, forecast-based financing initiatives, such as those of the Red Cross, aim to provide financial assistance to communities in advance of floods to enable an effective response (de Perez et al. 2015; Lopez et al. 2020).

Generally, warning systems are longstanding success stories with respect to protecting human life. Several low- and middle-income countries have made spectacular progress in reducing their mortality risk via warning systems in the last three decades (UN 2015). However, while early warning systems and timely evacuations have led to reduced loss of life, monetary losses have continued to grow [Centre for Research on the Epidemiology of Disasters (CRED); CRED and UNISDR 2018].

Important factors that may influence the effectiveness of flood early warning systems in reducing monetary losses are the lead time, the flood intensity, and the ability of civil protection and affected parties to undertake emergency measures effectively (Molinari and Handmer 2011; Morss et al. 2016). However, empirical studies on the effectiveness of flood early warning in reducing monetary losses are rare (Morss et al. 2016; Kreibich et al. 2017; Rai et al. 2020). Pappenberger et al. (2015) conduct an analysis for the European Flood Awareness System (EFAS), finding that every Euro invested in EFAS pays off, with a cost–benefit ratio between 1:4 and 1:409. Their sensitivity analysis highlights that the greatest uncertainty in these estimates comes from the avoided monetary losses, which reflect the wide range of possible responses to flood warnings.

The objective of this study is to gain more knowledge on the determinants of effective flood early warning with respect to reducing monetary losses. We analyze warning reception and monetary losses at the household property level, based on 4,468 loss cases from six floods in Germany since 2002 collected by four similar surveys, following a repeated cross-sectional design. We use propensity score matching (PSM) to calculate causal estimates of monetary loss reduction for different early warning receipt situations and a logit regression model to reveal which factors may be associated with knowing what to do when receiving a flood warning.

Data and methods

Study design. The study is composed of two parts: first, we quantify the average treatment effect of different flood early warning situations with respect to reducing building and contents loss, applying PSM with five matching variants (see “PSM” section). PSM is a bias-reduction technique that aims at providing causal estimates from observational data. The following

three treatments are used to indicate different “qualities of the treatment” represented by flood early warning situations:

- 1) At least 1-hour (h) warning lead time
- 2) At least 1-h warning lead time and people received a warning containing helpful information
- 3) At least 1-h warning lead time and people (stated that they) knew what to do

Second, we analyze with a logit regression model which factors are associated with people being more likely to know what to do when they receive a flood warning, since this was identified to be decisive for an effective reduction of monetary loss (see “Logit regression model” section).

Empirical flood loss data. The database for the analyses consists of data collected via computer-aided telephone interviews with private households that had experienced losses due to a flood between 2002 and 2013 in Germany (Table 1). On the basis of flood reports, press releases, and flood maps derived from satellite data (e.g., www.zki.dlr.de), lists of inundated streets were compiled separately after one or two flood events. Based on these lists, property-specific random samples of potentially affected households were generated, i.e., their telephone numbers were selected from the public telephone directory. The surveys were undertaken by professional survey institutes. At the beginning of the interview, each household was asked whether it had suffered monetary losses due to the specific flood event(s); the interview was only continued if this was the case. Additionally, the person on the phone was asked to identify the member of the household that had the best knowledge about the flood event and the incurred monetary losses. The interview was then conducted with this person, which lasted on average 30 min. The standardized questionnaires for all the survey campaigns contained about 180 questions including aspects related to hazard (e.g., inundation depth and duration, flow velocity), flood experience and awareness, early warning, emergency and precautionary measures, building and socioeconomic characteristics, and building and contents losses. Most questions were asked in a closed format, i.e., a list of possible answers was given (with either a single answer or multiple answers possible).

For instance, with respect to early warning, people were asked, “How did you become aware of the imminent flood hazard for you?” A list of possible answers was provided with multiple answers (and open answer) possible: storm warning, flood warning by public authority; call for evacuation; warning by neighbors, friends, relatives, or similar; general transregional media coverage; and the individual’s own research or observations. Thus, flood warning was treated in a broad sense, including official and unofficial warnings and even people’s own observations. Similarly, respondents were asked about the information content of the warning, with a

Table 1. Cross-sectional flood surveys: computer-aided telephone interviews with private households suffering flood loss (adapted from Sairam et al. 2019).

Characteristics	Surveys			
Date of survey	April/May 2003	November/December 2006	February/March 2012	February/March 2014
Flood(s)	August 2002	August 2005, April 2006	August 2010, January 2011	June 2013
Affected regions	Elbe and Danube catchments	Elbe and Danube catchments	Elbe, Oder, and Rhine catchments	Elbe, Danube, Rhine, and Weser catchments
No. of households interviewed	1,697	461	658	1,652
Survey response rate	15%	18%	16%	17%
References	Thieken et al. (2007)	Kreibich et al. (2011a)	Kienzler et al. (2015)	Kreibich et al. (2017)

checklist referring to different types of information. The questionnaire also included a question designed to reveal the perception of the interviewee regarding whether or not they knew what to do when the warning reached them. People were asked to assess this qualitatively on a scale from “1—it was completely clear to me” to “6—it was completely unclear to me.”

Cross-checks of answers during the interview were undertaken to improve data quality, since it allowed clarification of contradictory answers. Questions of the survey related to early warning are provided in appendix A (translated into English), the complete questionnaire (as used in 2006) is available online in German: www.gfz-potsdam.de/fileadmin/gfz/sec44/html/Questions_MEDIS.htm (last accessed 1 February 2021). These data and parts of them have been used before, e.g., to analyze what motivates households to undertake private precautionary measures (Kreibich et al. 2011a; Hudson et al. 2017), to quantify the effectiveness of these rather long-term measures (Hudson et al. 2014; Sairam et al. 2019) and how these jointly with other variables determine the amount of flood damage (Thieken et al. 2005; Merz et al. 2013).

Variables from 4,468 interviews with flood-affected private households are available for this analysis as described in Table 2. Some variables are direct answers to a question, like the warning lead time. Additionally, indicators were developed. For instance, the indicator of flood warning information is the sum of the assessment points for the single pieces of information contained in the warning (Thieken et al. 2005). The more helpful an information appeared, the more assessment points were assigned, based on expert judgment: behavioral tips and recommendations for self-protection—4 points; information about the storm: time of occurrence, endangered region—2 points; information about the storm: expected amount of rainfall—2 points; information about the flood: gauge height—2 points; information about the flood: areas at risk—2 points; information about levee or dam breaches—2 points; information about evacuations—1 point; information about detours or road closures—1 point (Thieken et al. 2005). Details how the indicators of flood warning source, precautionary measures, flood experience and socioeconomic status were calculated are also given in Thieken et al. (2005).

Since not all people were willing to answer all questions, not all variables are available for each damage case. For instance, there are 2,283 observations for the building loss ratio and 2,705 for the household contents loss ratio (Table 2). These building and contents loss ratios were calculated consistently for all surveys as follows: the absolute building and contents losses reported by the surveyed households were divided by the building and contents values, respectively, at the time of the flood event. Actuarial valuation method VdS guideline 772 1988-10 (Dietz 1999), which is commonly used in the insurance sector for Germany, was used to estimate absolute values of residential building in terms of replacement costs. Also, the value of household contents was estimated following an approach from the insurance sector, which is a common approach in Germany. This is done by multiplying the living area of the residential building or flat with the average contents value per square meter. For 2005, an average content value of 650 EUR m⁻² was used; this value was adapted to the other flood event years using a comparison of the relative consumer price indices to account for inflation.

Propensity score matching. PSM is a technique for reducing bias in empirical analysis, with the overall objective to produce causal estimates of a treatment effect when nonexperimental survey data are analyzed (Rosenbaum and Rubin 1983; Hudson et al. 2014). Comparing the average outcome of the treatment group with that of the nontreatment group could provide an estimate of the effectiveness of the treatment if people were randomly allocated to the groups. However, this is likely not the case in observational data (e.g., a survey with voluntary participation). Rosenbaum and Rubin (1983) noted that through conditioning on the confounders (i.e., the variables most likely to lead to selection bias due to their influence on both the outcome and participation in the treatment group) it may be possible to find survey

Table 2. Description of the variables and their use in the statistical analyses (adapted from Hudson et al. 2014; Sairam et al. 2019).

Variable	Type* and range	Number of observations**	Statistical analyses	Use of variable
Building loss ratio (brloss)	C: 0 (no loss) to 1 (total loss)	2,283	PSM	Outcome variable measuring treatment effect
Contents loss ratio (crloss)	C: 0 (no loss) to 1 (total loss)	2,705	PSM	Outcome variable measuring treatment effect
Absolute building loss (bloss)	C: 15 to 1,000,000 EUR	2,671	PSM	Outcome variable measuring treatment effect
Absolute contents loss (closs)	C: 0 to 400,000 EUR	2,784	PSM	Outcome variable measuring treatment effect
Warning lead time (wt)	C: 0 to 998 h	2,774	PSM	Treatment variable
Indicator of flood warning information (wi)	O: 0 = no helpful information to 16 = a lot of helpful information	4,191	PSM	Treatment variable
Knowledge of how to protect oneself (wq)	O: 1 = receiver of warning knew exactly what to do to 6 = receiver of warning had no idea what to do	2,994	PSM Logit regression model	Treatment variable Dependent variable to indicate potential links with treatment group
Indicator of flood warning source (ws)	O: 0 = no warning to 4 = official warning through authorities	4,427	Logit regression model	Potential predictor variable
Precautionary measures indicator (pre)	O: 0 = no measures undertaken to 2 = efficient measures undertaken	4,468	PSM Logit regression model	Confounder variable of outcome and treatment participation Potential predictor variable of wq
Perception of efficiency of private precaution (epre)	O: 1 = very efficient to 6 = not efficient at all	4,185	Logit regression model	Potential predictor variable of wq
Flood experience indicator (fe)	O: 0 = no experience to 9 = recent flood experience	4,236	PSM Logit regression model	Confounder variable of outcome and treatment participation Potential predictor variable of wq
Knowledge of flood hazard (kh)	N: (yes/no)	4,363	Logit regression model	Potential predictor variable of wq
Age of the interviewed person (age)	C: 16 to 99 yr	4,213	Logit regression model	Potential predictor variable of wq
Socioeconomic status (socp)	O: 3 = very low socioeconomic status to 13 = very high socioeconomic status	3,129	PSM	Confounder variable of outcome and treatment participation
Survey	N: 1—survey conducted in 2002; 2—survey conducted between 2005 and 2011, 3—survey conducted in 2013	4,468	PSM	Confounder variable of outcome and treatment participation
Return period (rp)	C: 1 to 6,685 yr	3,383	PSM	Confounder variable of outcome and treatment participation
Household size (hs)	C: 1 to 20 people	4,326	PSM	Confounder variable of outcome and treatment participation
Water depth (wst)	C: 248 cm below ground to 1,328 cm above ground	4,275	PSM	Confounder variable of outcome and treatment participation
Duration (d)	C: 1 to 1,440 h	4,256	PSM	Confounder variable of outcome and treatment participation
Number of children (<14 yr) in household (chi)	C: 0 to 6	3,648	Logit regression model	Potential predictor variable of wq
Number of elderly (>65 yr) in household (eld)	C: 0 to 9	3,757	Logit regression model	Potential predictor variable of wq
Ownership structure (own)	N (1 = tenant; 2 = owner of flat; 3 = owner of building)	4,467	Logit regression model	Potential predictor variable of wq
Monthly net income (inc)	O: 11 = below 500 EUR to 16 = 3,000 EUR and more	3,098	Logit regression model	Potential predictor variable of wq

* Scaling of variables C: continuous, O: ordinal, N: nominal.

** Since not all people were willing to answer all questions, not all information is available for each interview.

respondents who are similar enough to act as counterfactual observations and remove this bias. This would help to produce more reliable estimates. PSM aims to achieve this by converting the information contained in all relevant confounder variables into the single propensity score which can be used as a holistic indicator of the overall similarity of two respondents. Assuming certain conditions are met (see appendix B), an observation in the nontreatment group that has a propensity score that is the same (or sufficiently close) in value to that of an observation in the treatment group allows for us to judge the benefit of the treatment by comparing the differences in the outcomes of these two observations. A full list of the confounding variables is presented in Table 2, in which continuous variables were left unchanged. Ordinal and nominal variables were converted into dummy variables for the relevant categories. These variables were selected following the approaches developed in Hudson et al. (2014) and Sairam et al. (2019). Hudson et al. (2014) suggest that using multiple matching methods can help provide an overall indication of the reliability of the results by observing the spread of the results. Hence, we apply five matching methods, i.e., nearest-neighbor matching, Epanechnikov kernel matching, Gaussian kernel matching, radius matching, and stratification matching. Two main outcome variables are considered: the building loss ratio and the contents loss ratio. However, to support the interpretation, the average treatment effect of the treatment that results in significant loss mitigation is additionally quantified in terms of absolute building and contents loss. See appendix B for more explanation and details of PSM and its application in this study.

Logit regression model. The analysis is undertaken to explore the factors associated with respondents knowing what to do in case they receive a flood warning. The analysis is conducted by employing a backward stepwise variable removal process (Wooldridge 2002; Fields 2009). Therefore, an initial logit regression model is employed, see Eq. (1), in which the probability of knowing what to do for individual i is a function of a constant term (α), β are vectors of coefficient terms, \mathbf{H}_i is a vector of hydrological/flood-related factors as a proxy of risk, \mathbf{EM}_i is a vector of emergency measures and response factors, \mathbf{FE} is a vector of flood-related experiences, and \mathbf{SES}_i is a vector of socioeconomic status factors, while ε_i represents the error term. A logit regression model is employed because the dependent variable (knowing what to do) is binary (i.e., 0 or 1). A full list of variables is presented in Table 2 in which continuous variables were left unchanged and ordinal and nominal variables were converted into dummy variables for the relevant categories:

$$\text{logit}[p(\text{know})_i] = \alpha + \beta_1 \mathbf{H}_i + \beta_2 \mathbf{EM}_i + \beta_3 \mathbf{FE} + \beta_4 \mathbf{SES}_i + \varepsilon_i \quad (1)$$

Once the logit regression model is estimated, the least statistically significant variable, judged by p values corresponding to a t test, is removed and the revised logit regression model is reestimated. This process is repeated until only variables that are statistically significant at the 5% level remain. The logit regression model is nonlinear, which makes it difficult to interpret the coefficient estimates outside of being positive or negative. Therefore, we also provide the average marginal effect estimates. These values can be understood as the change in probability when the model is evaluated at the sample average values (e.g., average age or flood experience), just as the regression coefficients can be interpreted in a linear regression model.

This approach has been used in a range of studies within natural hazard or climate risk research (Bubeck et al. 2013; Poussin et al. 2014; Sarmiento et al. 2020). One caveat to keep in mind, however, is that it is exploratory in nature as it examines the relative strength of correlations rather than the structure of the relationships. However, this approach is a useful start and it could be complemented with longitudinal and sociopsychological datasets for further in-depth studies of causal linkages (Hudson et al. 2019).

Results

Treatment effect of flood early warning. Flood warning is only effective in reducing monetary losses of residential buildings and contents when people know what to do when they receive the warning (Fig. 1). In detail, PSM with five different matching methods reveals the following: Receiving a warning with a lead time of at least 1 h, even when this warning contains helpful information, does not lead to a significant effect on the building and contents loss ratios, not in terms of statistical significance nor in size of the impact. Only when people know what to do when the warning reaches them (with at least 1 h of lead time) can a significant loss reduction be achieved. The average reduction of the household contents loss ratio is 4 percentage points (averaged across all matching methods), a reduction of 3,800 EUR for the average treatment recipient (Fig. 1, averaged across all matching methods). This is substantial in comparison with the mean (median) contents loss ratio of 21% (10%) and absolute contents loss of 17,000 (7,700) EUR. For the building loss ratio, the average reduction is 2 percentage points (averaged across all matching methods), a loss reduction of 10,000 EUR (Fig. 1, averaged across all matching methods). This is a remarkable reduction in comparison with the mean (median) building loss ratio of 11% (5%) and absolute building loss of 48,000 (22,000) EUR. These average loss reductions are lower but still considerable in comparison with reported average loss reductions of up to 15,000 EUR due to long-term precautions taken privately by individuals (Hudson et al. 2014; Sairam et al. 2019). However, building precautionary measures like sealing the cellar or flood proofing the heating and oil tank are quite costly to install (Kreibich et al. 2011b). A point to consider is that most of the sample observations originate from the two larger flood events in 2002 and 2013. Thus, the loss reduction values more closely represent the outcome in rather extreme scenarios.

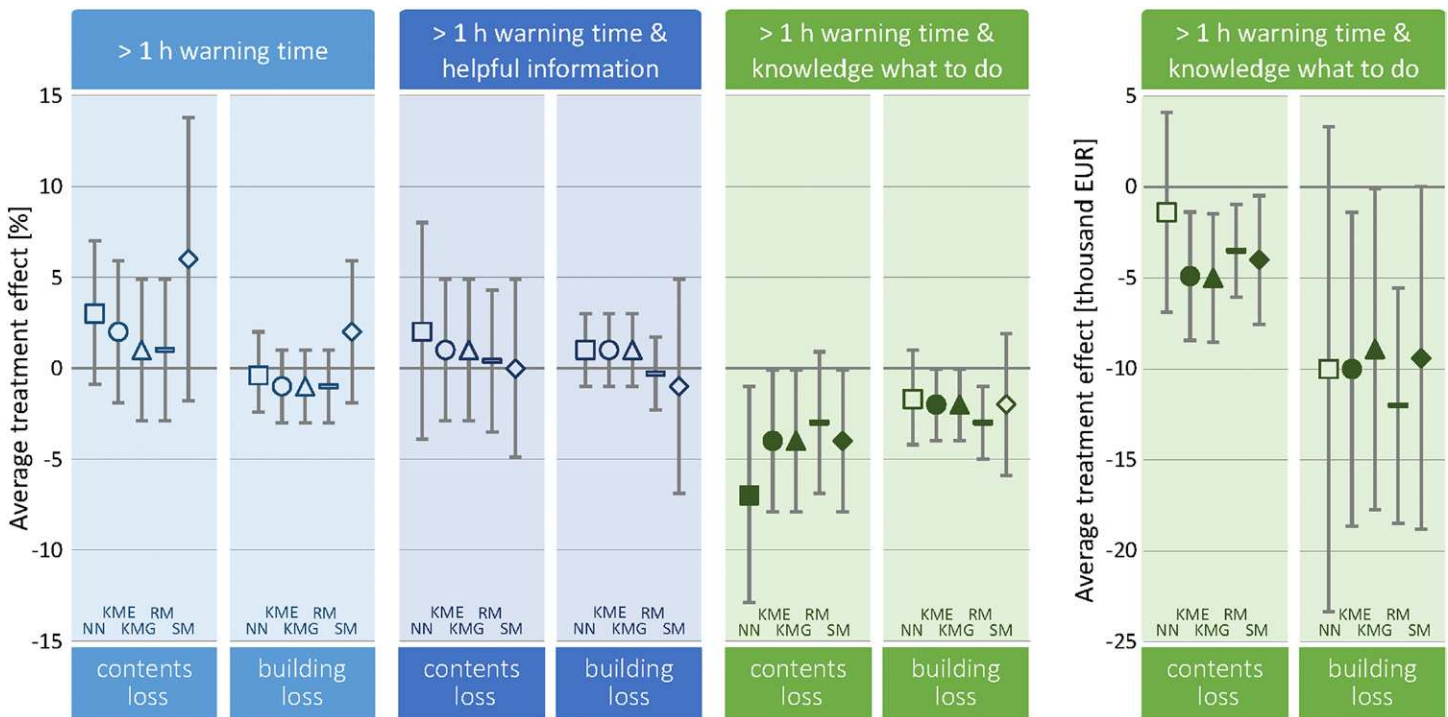


Fig. 1. Average treatment effects estimated with nearest-neighbor matching (NN), kernel matching (Epanechnikov) (KME), kernel matching (Gaussian) (KMG), radius matching (RM), and stratification matching (SM). Analyzed are the effects of the treatments: “at least 1-h warning lead time only,” “at least 1-h warning lead time and helpful warning information,” and “at least 1-h warning lead time and knowledge of what to do” for relative (and absolute) building and contents loss (error bars show 95% confidence intervals). Negative treatment effects indicate loss reduction. Filled symbols indicate statistically significant results ($p < 0.05$; for details; see appendix C).

Additionally, this analysis reveals that it is potentially easier to reduce contents loss than building loss, e.g., all the contents loss related estimates across the different matching methods are significant, while only three out of five estimates for building loss are statistically significant (Fig. 1). This may be due to more emergency measures available for limiting loss to household contents than to buildings, for instance as expensive objects can be moved to higher floors (Kreibich et al. 2017). However, if effective, protecting the building (e.g., via preventing inflowing water using shutters) saves more money on average.

With respect to external validity, i.e., transferability of these results to other areas, it needs to be noted, that the results are based on data of several floods in Germany. Thus, they are valid for Germany and similar countries. For instance, the statement “I was not capable of doing anything” plays a minor role when people are asked why they did not undertake emergency measures (Kienzler et al. 2015) and response costs and financial considerations have a minor influence on flood mitigation behavior in Germany (Bubeck et al. 2013). This may be different in developing countries, for example, where residents themselves may not have the capacity to take effective response measures. In such areas, flood damage reduction may require more support, e.g., timely humanitarian action, to overcome potential barriers that prevent warned people from putting their knowledge into action.

Knowing what to do. Since it is decisive for loss reduction via early warning, we further analyze with a logit regression model that factors are associated with people knowing what to do when they receive a flood warning (Table 3). Our analysis shows that people who have undertaken precautionary measures, have flood experience, and are supported by helpful warning information are more likely to know what to do when they receive a flood warning. This was revealed through the high marginal effects of these variables as linked to possessing

Table 3. Results of the variables associated with possessing sufficient knowledge regarding what to do when a flood warning is received. Standard errors in parentheses: three asterisks (*) indicate $p < 0.01$, two asterisks (**) indicate $p < 0.05$, and one asterisk (*) indicates $p < 0.1$.**

	Parameter estimate	Marginal effect
Age of the interviewed person	0.006** (0.003)	0.001** (0.0006)
Knowledge of flood hazard	0.73*** (0.12)	0.16*** (0.02)
Flood experience indicator	0.86*** (0.09)	0.18*** (0.02)
Perception of efficiency of private precautions	0.51*** (0.08)	0.11*** (0.02)
Some precautionary measures undertaken	0.8*** (0.09)	0.18*** (0.02)
Many precautionary measures undertaken	1.26*** (0.11)	0.29*** (0.02)
Indicator of flood warning information	1.35*** (0.1)	0.27*** (0.02)
Official warning through authorities	0.4*** (0.09)	0.09*** (0.02)
Constant	-3.72*** (0.19)	
Observations		4,161

sufficient knowledge about what to do (Table 3). The most powerful association, with a marginal effect of an increase of 29 percentage points in the likelihood of knowing what to do, is having undertaken many precautionary measures before. Thus, we quantitatively show that good, long-term preparedness is helpful in various ways; besides loss mitigation due to precautionary measures (Hudson et al. 2014; Sairam et al. 2019), it also supports loss mitigation due to effective early warning and emergency response. This is in accordance with previous research, which found that people who are proactive in one area of flood risk management are proactive, and thus more effective, in other areas as well (Osberghaus 2015; Hudson et al. 2017).

Conclusions

We quantitatively show that a significant monetary loss reduction is only achieved when people know what to do when they receive a timely flood warning. Hence, our study provides quantitative evidence for the calls of early warning experts to add helpful information to warning messages and improving emergency communication [e.g., Grundfest et al. 1978; World Meteorological Organization (WMO); WMO 2019]. Additionally, we show that next to the support by helpful warning information, people who have undertaken precautionary measures and have flood experience are more likely to know what to do when they receive a flood warning. Thus, effective risk communication, training and (financial) support for private precaution are helpful twice; in addition to loss mitigation through precautionary measures comes loss mitigation due to more effective emergency response.

Acknowledgments. The surveys collecting the empirical flood loss data in Germany were supported by the German Research Network Natural Disasters [German Ministry of Education and Research (BMBF), 01SFR9969/5], by the MEDIS project (BMBF; 0330688), by the project “Hochwasser 2013” (BMBF; 13N13017), and by a joint venture between the German Research Centre for Geosciences GFZ, the University of Potsdam, and the Deutsche Ruckversicherung AG, Duesseldorf.

Data availability statement. Flood loss data were partly funded by a joint venture between the German Research Centre for Geosciences GFZ, the University of Potsdam, and the reinsurance company Deutsche Ruckversicherung AG (www.deutscherueck.de) and is, as such, proprietary and not publicly accessible, but may be obtained upon request.

Appendix A: Early warning–related questions of the surveys

The following list of questions (Q) and answers (A) of the survey is translated from German to English. The complete questionnaire as used in 2006 is available online in German: www.gfz-potsdam.de/fileadmin/gfz/sec44/html/Questions_MEDIS.htm (last accessed 1 February 2021).

Q: Please think back to the days before the event. How did you become aware of the imminent flood hazard for you? (multiple answers possible)

A: Storm warning, e.g., on the radio, TV, Internet, by SMS, etc.

Flood warning by authority or disaster control (e.g., fire department, police)

Call for evacuation

Warnings by neighbors, friends, relatives, or similar

General trans-regional media coverage

Own research, e.g., tide gauge information, tide gauge level

Own observation

Other warning, namely (record openly):

Did not become aware of the danger at all/ was not warned/was surprised by the flood

Do not know

Not specified

Q: How many hours before the flooding occurred did the warning reach you or did you become aware of the danger yourself? (If there were several warnings, the earliest is meant)

A: Indication in hours:

Indication in days:

Do not know

Not specified

Q: Which of the following information was included in the warning?

A: Information about the storm: time of occurrence, endangered region

Information about the storm: expected amount of rainfall

Information about the flood: gauge height, i.e., time and/or height of the maximum water level

Information about the flood: areas at risk

Behavioral tips and recommendations for self-protection (e.g., turn off electricity, move inventory to higher floors, lock windows and doors, employ sandbags, move motor vehicles)

Information about evacuations

Information about levee or dam breaches

Information about detours or road closures

Other information, namely (record openly):

None of this information

Do not know

Not specified

Q: Did you know how to protect yourself and your household from flooding before the flood threat became acute for you? Please give me a number between 1 for 'it was completely clear to me' and 6 for 'it was completely unclear to me'. You can use the values in between to grade your answer.

A: (1) it was completely clear to me

(2)

(3)

(4)

(5)

(6) it was completely unclear to me

Do not know

Not specified

Appendix B: Application of PSM

PSM was developed to evaluate an intervention's success by estimating the average treatment effect on the treated (ATT) as defined in Eq. (1). Below, $E(\cdot)$ is the expectations operator, T is a binary variable for participation in the treatment group or not, y_1 is the outcome under treatment, and y_0 is the outcome under nontreatment. Therefore, it can be seen that the ATT is the change due to the intervention:

$$ATT = E(y_1 - y_0 | T = 1). \quad (B1)$$

However, estimating Eq. (B1) is difficult, as either the outcome under treatment [$E(y_1 | T = 1)$] or the outcome under nontreatment [$E(y_0 | T = 0)$] is observed. Therefore, the natural method for estimating the ATT in this case is presented in Eq. (B2), which is using the difference between sample subgroup averages. However, Angrist and Piske (2009) show that this choice could result in the estimated ATT (\widehat{ATT}) being a combination of the

ATT and selection bias (SB). This is because of the absence of a suitable counterfactual observation due to the nonexperimental data generation process that observational or survey data entail:

$$\widehat{ATT} = E(y_1 | T=1) - E(y_0 | T=0) = ATT + SB. \quad (B2)$$

Rosenbaum and Rubin (1983) noted that through conditioning on the confounders (i.e., the variables most likely to lead to $SB \neq 0$ due to their influence on both the outcome and participation in the treatment group) it may be possible to find survey respondents who are similar enough to act as counterfactual observations. This would help limit selection bias and produce more reliable estimates. However, if the raw confounder variable values are directly used to find comparable respondents, a dimensionality issue (i.e., too many factors seeking exact comparisons at the same time) may occur. This is because it is unlikely that two observations will be sufficiently similar with regard to all relevant variable values. PSM mitigates this dimensionality issue by converting the information contained in all relevant confounder variables into a single score that can be used as a holistic indicator of the overall similarity of two respondents. PSM is able to achieve this through conditioning on the propensity score (PS), which is the probability of participating in the “treatment group” (e.g., receiving a flood warning with at least one hour warning lead time) as estimated in relation to the important confounding variables. The confounding variables are those that explain both outcome and treatment. The actual estimated probability itself, or its accuracy in explaining participation, is relatively unimportant, since its purpose is to collapse the relevant information explaining both participation and outcome into a single value (Rosenbaum 2002).

Therefore, PSM is a two-stage process. The first stage estimates the probability of participation in the treatment group in relation to the key confounders, often estimated via a logit or probit model. The second stage then compares the average difference between sufficiently comparable observations by matching respondents who have suitable PS values. There are several ways to produce matches (Hudson et al. 2014). In order for the PSM approach to move toward unbiased or causal estimates, the following three conditions need to hold:

- 1) Unconfoundedness: $(y_1, y_0) \perp T | p(X)$
- 2) Balancing: $T \perp X | p(X)$
- 3) Overlap: The probability distributions for the control and treatment groups share a common support (i.e., overlapping probability ranges)

Condition 1 implies that treatment and potential outcomes are independent of one another, conditional on the PS, which allows for bias reduction to occur by causing SB in Eq. (B2) to shrink. Condition 2 is that, when conditioned on $p(X)$, treatment participation and individual traits are independent of one another. When condition 2 holds, the PS is a balancing score, and then matching on the value of the PS achieves the same as conditioning on raw confounder values. Condition 3 implies that the observations have a similar enough PS to create a sample of good matches. Of these three conditions, the unconfoundedness cannot be tested and must be assumed to hold (based on the theory-driven approach to variable selection). The balancing condition can be tested via a series of t tests. We created automated strata of observations where the mean PS was the same for the treatment and control group observations. Then, for each variable within a given strata it was tested via a t test if it differed significantly between the treatment and control group observations. This approach allowed us to approximate conditioning on the PS while

providing enough observations to conduct statistical tests. For our models, the balancing condition was found to hold. The overlap condition can be investigated by looking at the estimated PSs to compare the ranges of values. A common action is to restrict observations to values that fall under the common area underneath the probability curves, known as the common support. We restricted our sample to this area to maximize the comparability of potential matches.

Additionally, Hudson et al. (2014) suggest that using multiple matching methods can help provide an overall indication of the reliability of the results. This is because the better the three conditions hold, the more likely it is that different matching methods will produce ATT estimates that are close together in value.

Therefore, a key element is the selection of the confounding variables used to generate the PS values that result in condition 1 being met. However, condition 1 cannot be formally tested (a limitation of PSM and observational data more generally). Therefore, confounder selection is driven by the need to select the relevant set of variables that produces the highest likelihood of condition 1 holding.

To achieve this outcome, we adapt the variable selection approach developed in Hudson et al. (2014). From the list of flood loss determining variables identified by Merz et al. (2013), we select those variables that, based on expert judgment, are most likely to jointly explain both the flood loss suffered and the likelihood of participation in one of the treatment groups, e.g., receiving a flood warning with at least one hour lead time. This approach to variable selection was also successfully employed in Sairam et al. (2019). Additionally, in order to improve the balancing condition, variables known to explain only the outcome can be included as a secondary set of variables (Hudson et al. 2014). The confounding variables that have been selected are presented in Table 2.

There are two main outcome variables considered for the PSM: the building loss ratio and the contents loss ratio. Relative values are selected because they result in better behaved outcome variables in comparison with the absolute losses. However, to support the interpretation of treatment 3, which results in significant loss mitigation effects, its ATT is additionally quantified in terms of absolute building and contents loss. This is to provide an additional nuance to the relative effect, by understanding the size of the impact in a more pragmatic sense.

One possible limitation of the PSM analysis is that it can be perceived that treatments 2 and 3, as compared to treatment 1, are not fully explained by the selected confounders due to their more sophisticated nature. Therefore, some selection bias could still be present in the estimated ATT values presented in Table C1. However, following the theoretical baseline set out in Hudson et al. (2014), the most important factors are likely to have been accounted for. This can also be seen through its successful implementation in Sairam et al. (2019).

Appendix C: Detailed statistical results

This appendix presents a more detailed view of the statistical results presented in the figures and text of the main manuscript. Tables C1 and C2 present PSM results for the relationship between the treatments with respect to flood warning and loss ratios and the PSM results for the relationship between treatment 3 and absolute building and contents loss.

Table C1. PSM results for the relationship between the treatments with respect to flood warning and loss ratios. Standard errors in parentheses are calculated via bootstrapping with 1,000 repetitions; spread of ATT is calculated as the ratio of the standard deviation of the estimates to their mean value (also known as “coefficient of variation”). Three asterisks (*) indicates $p < 0.01$, two asterisks (**) indicate $p < 0.05$, and one asterisk (*) indicates $p < 0.1$.**

Treatment type	Contents loss ratio ATT (standard error)			Building loss ratio ATT (standard error)		
	1	2	3	1	2	3
	At least 1-h warning lead time ($w_t \geq 1$)	At least 1-h warning lead time ($w_t \geq 1$) and warning contained helpful information ($w_i \geq 7$)	At least 1-h warning lead time ($w_t \geq 1$) and people knew what to do ($w_q < 3$)	At least 1-h warning lead time ($w_t \geq 1$)	At least 1-h warning lead time ($w_t \geq 1$) and warning contained helpful information ($w_i \geq 7$)	At least 1-h warning lead time ($w_t \geq 1$) and people knew what to do ($w_q < 3$)
Nearest-neighbor matching	0.03 (0.02)	0.02 (0.03)	-0.07** (0.03)	-0.004 (0.01)	0.01 (0.01)	-0.017 (0.013)
Kernel matching method (Epanechnikov)	0.02 (0.02)	0.01 (0.02)	-0.04** (0.02)	-0.01 (0.01)	0.01 (0.01)	-0.02** (0.01)
Kernel matching method (Gaussian)	0.01 (0.02)	0.01 (0.02)	-0.04** (0.02)	-0.01 (0.01)	0.01 (0.01)	-0.02** (0.01)
Radius matching method	0.01 (0.02)	0.004 (0.02)	-0.03** (0.15)	-0.01 (0.01)	-0.003 (0.01)	-0.03*** (0.01)
Stratification matching	0.06 (0.04)	-0.0004 (0.025)	-0.04** (0.02)	0.02 (0.02)	-0.01 (0.03)	-0.02 (0.02)
Average ATT estimate	0.03	0.01	-0.04	-0.00	0.00	-0.02
Spread of ATT estimate (absolute value)	0.8	0.89	0.34	4.67	2.76	0.23
Effective	No	No	Yes	No	No	Yes

Table C2. PSM results for the relationship between treatment 3 and absolute building and contents loss rounded to two significant figures. Standard errors in parentheses as calculated via bootstrapping with 1,000 repetitions; spread of ATT is calculated as the ratio of the standard deviation of the estimates to their mean value (also known as “coefficient of variation”). Three asterisks (*) indicate $p < 0.01$, two asterisks (**) $p < 0.05$, and one asterisk (*) indicates $p < 0.1$.**

Treatment type	Contents loss ATT (standard error) in EUR	Building loss ATT (standard error) in EUR
	3) At least 1-h warning lead time ($w_t \geq 1$) and people knew what to do ($w_q < 3$)	3) At least 1-h warning lead time ($w_t \geq 1$) and people knew what to do ($w_q < 3$)
Nearest-neighbor matching	-1,400 (2,800)	-10,000 (6,800)
Kernel matching method (Epanechnikov)	-4,900*** (1,800)	-10,000*** (4,400)
Kernel matching method (Gaussian)	-5,000*** (1,800)	-8,900*** (4,500)
Radius matching method	-3,500*** (1,300)	-12,000*** (3,300)
Stratification matching	-4,000** (1,800)	-9,400* (4,800)
Average ATT estimate	-3,800	-10,000
Spread of ATT estimate (absolute value)	0.39	0.11
Effective	Yes	Yes

References

- Angrist, J., and J. Piske, 2009: *Mostly Harmless Econometrics*. Princeton University Press, 373 pp.
- Bubeck, P., W. J. W. Botzen, H. Kreibich, and J. C. J. H. Aerts, 2013: Detailed insights into the influence of flood-coping appraisals on mitigation behavior. *Global Environ. Change*, **23**, 1327–1338, <https://doi.org/10.1016/j.gloenvcha.2013.05.009>.
- CRED and UNISDR, 2018: Economic losses, poverty & disasters: 1998–2017. CRED and UNISDR Rep., 33 pp., www.preventionweb.net/files/61119_credeconomiclosses.pdf.
- de Perez, E. C., B. van den Hurk, M. K. van Aalst, B. Jongman, T. Klose, and P. Suarez, 2015: Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Nat. Hazards Earth Syst. Sci.*, **15**, 895–904, <https://doi.org/10.5194/nhess-15-895-2015>.
- Deutsch, M., and K.-H. Pörtge, 2001: Die Hochwassermeldeordnung von 1889—Ein Beitrag zur Geschichte des Hochwasserwarn und Meldedienstes in Mitteldeutschland. *Forum Katastrophenvorsorge 2001*, Leipzig, Germany, DKKV, 396–405.
- Dietz, H., 1999: *Wohngebäudeversicherung: Kommentar*. 2nd ed. VVW Verlag Versicherungswirtschaft GmbH, 756 pp.
- Fields, A., 2009: *Discovering Statistics Using SPSS*. SAGE, 821 pp.
- Grundfest, E., G. F. White, T. C. Downing, 1978: Big Thompson flood exposes need for better flood reaction system to save lives. *Civil Eng.*, **48**, 72–73.
- Hudson, P., W. Botzen, H. Kreibich, P. Bubeck, and J. C. J. H. Aerts, 2014: Evaluating the effectiveness of flood damage mitigation measures by the application of propensity score matching. *Nat. Hazards Earth Syst. Sci.*, **14**, 1731–1747, <https://doi.org/10.5194/nhess-14-1731-2014>.
- , ——, J. Czajkowski, and H. Kreibich, 2017: Moral hazard in natural disaster insurance markets: Empirical evidence from Germany and the United States. *Land Econ.*, **93**, 179–208, <https://doi.org/10.3368/le.93.2.179>.
- , A. H. Thieken, and P. Bubeck, 2019: The challenges of longitudinal surveys in the flood risk domain. *J. Risk Res.*, **23**, 642–663, <https://doi.org/10.1080/13669877.2019.1617339>.
- Kienzler, S., I. Pech, H. Kreibich, M. Müller, and A. H. Thieken, 2015: After the extreme flood in 2002: Changes in preparedness, response and recovery of flood-affected residents in Germany between 2005 and 2011. *Nat. Hazards Earth Syst. Sci.*, **15**, 505–526, <https://doi.org/10.5194/nhess-15-505-2015>.
- Kreibich, H., S. Christenberger, and R. Schwarze, 2011a: Economic motivation of households to undertake private precautionary measures against floods. *Nat. Hazards Earth Syst. Sci.*, **11**, 309–321, <https://doi.org/10.5194/nhess-11-309-2011>.
- , I. Seifert, A. H. Thieken, E. Lindquist, K. Wagner, and B. Merz, 2011b: Recent changes in flood preparedness of private households and businesses in Germany. *Reg. Environ. Change*, **11**, 59–71, <https://doi.org/10.1007/s10113-010-0119-3>.
- , M. Müller, K. Schröter, and A. H. Thieken, 2017: New insights into flood warning reception and emergency response by affected parties. *Nat. Hazards Earth Syst. Sci.*, **17**, 2075–2092, <https://doi.org/10.5194/nhess-17-2075-2017>.
- Lopez, A., E. C. de Perez, J. Bazo, P. Suarez, B. van den Hurk, and M. van Aalst, 2020: Bridging forecast verification and humanitarian decisions: A valuation approach for setting up action-oriented early warnings. *Wea. Climate Extremes*, **27**, 100167, <https://doi.org/10.1016/j.wace.2018.03.006>.
- Merz, B., H. Kreibich, and U. Lall, 2013: Multi-variate flood damage assessment: A tree-based data-mining approach. *Nat. Hazards Earth Syst. Sci.*, **13**, 53–64, <https://doi.org/10.5194/nhess-13-53-2013>.
- Molinari, D., and J. Handmer, 2011: A behavioural model for quantifying flood warning effectiveness. *J. Flood Risk Manage.*, **4**, 23–32, <https://doi.org/10.1111/j.1753-318X.2010.01086.x>.
- Morss, R., K. J. Mulder, J. K. Lazo, and J. L. Demuth, 2016: How do people perceive, understand, and anticipate responding to flash flood risks and warnings? Results from a public survey in Boulder, Colorado, USA. *J. Hydrol.*, **541**, 649–664, <https://doi.org/10.1016/j.jhydrol.2015.11.047>.
- Osberghaus, D., 2015: The determinants of private flood mitigation measures in Germany—Evidence from a nationwide survey. *Ecol. Econ.*, **110**, 36–50, <https://doi.org/10.1016/j.ecolecon.2014.12.010>.
- Pappenberger, F., H. L. Cloke, D. J. Parker, F. Wetterhall, D. S. Richardson, and J. Thielen, 2015: The monetary benefit of early flood warnings in Europe. *Environ. Sci. Policy*, **51**, 278–291, <https://doi.org/10.1016/j.envsci.2015.04.016>.
- Perera, D., O. Seidou, J. Agnihotri, M. Rasmy, V. Smakhtin, P. Coulibaly, and H. Mehmood, 2019: Flood early warning systems: A review of benefits, challenges and prospects. UNU-INWEH Rep. Series 08, 30 pp., <https://inweh.unu.edu/wp-content/uploads/2019/08/Flood-Early-Warning-Systems-A-Review-Of-Benefits-Challenges-And-Prospects.pdf>.
- Poussin, J. K., W. J. W. Botzen, and J. C. J. H. Aerts, 2014: Factors of influence on flood damage mitigation behavior by households. *Environ. Sci. Policy*, **40**, 69–77, <https://doi.org/10.1016/j.envsci.2014.01.013>.
- Rai, R. K., M. J. C. van den Homberg, G. P. Ghimire, and C. McQuistan, 2020: Cost-benefit analysis of flood early warning system in the Karnali River basin of Nepal. *Int. J. Disaster Risk Reduct.*, **47**, 101534, <https://doi.org/10.1016/j.ijdrr.2020.101534>.
- Rosenbaum, P., 2002: *Observational Studies*. Springer, 375 pp.
- , and D. Rubin, 1983: The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55, <https://doi.org/10.1093/biomet/70.1.41>.
- Sairam, N., K. Schröter, S. Lüdtkke, B. Merz, and H. Kreibich, 2019: Quantifying flood vulnerability reduction via private precaution. *Earth's Future*, **7**, 235–249, <https://doi.org/10.1029/2018EF000994>.
- Sarmiento, J. P., V. Sandoval, and M. Jerath, 2020: The influence of land tenure and dwelling occupancy on disaster risk reduction. The case of eight informal settlements in six Latin American and Caribbean countries. *Prog. Disaster Sci.*, **5**, 100054, <https://doi.org/10.1016/j.pdisas.2019.100054>.
- Stephens, E., and H. Cloke, 2014: Improving flood forecasts for better flood preparedness in the UK (and beyond). *Geogr. J.*, **180**, 310–316, <https://doi.org/10.1111/geoj.12103>.
- Thieken, A. H., M. Müller, H. Kreibich, and B. Merz, 2005: Flood damage and influencing factors: New insights from the August 2002 flood in Germany. *Water Resour. Res.*, **41**, W12430, <https://doi.org/10.1029/2005WR004177>.
- , H. Kreibich, M. Müller, and B. Merz, 2007: Coping with floods: Preparedness, response and recovery of flood-affected residents in Germany in 2002. *Hydrol. Sci. J.*, **52**, 1016–1037, <https://doi.org/10.1623/hysj.52.5.1016>.
- UN, 2015: GAR15—Global Assessment Report on Disaster Risk Reduction 2015—Making development sustainable: The future of disaster risk management. UNISDR Rep., 310 pp., www.preventionweb.net/english/hyogo/gar/2015/en/home/GAR_2015/GAR_2015_1.html.
- UNDP, 2018: Five approaches to build functional early warning systems. UNDP Rep., 67 pp., www.eurasia.undp.org/content/dam/rbec/docs/UNDP%20Brochure%20Early%20Warning%20Systems.pdf.
- UNISDR, 2015: Sendai framework for disaster risk reduction 2015–2030. UNISDR Rep., 37 pp., www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf.
- WMO, 2019: Hazard warnings must reach the last mile, metre—and bucket. WMO, <https://public.wmo.int/en/media/news/hazard-warnings-must-reach-last-mile-metre-and-bucket>.
- Wooldridge, J. M., 2002: *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 1064 pp.