Developing and Pilot-Testing Societal Outcome Measures for the National Weather Service Weather-Ready Nation Initiative

Final Report

Submitted to:

National Weather Service

Silver Spring, MD

Submitted by:

Eastern Research Group, Inc. (ERG)

Lexington, MA

7/20/2016

CONTENTS

EXECUTIVE SUMMARY

This report summarizes the work that ERG performed to develop and pilot test a set of societal outcome metrics for the National Weather Service (NWS) Weather-Ready Nation (WRN) initiative. ERG's initial work on the project considered a subset of five WRN programs and products. Our research into these five programs and products, however, led to the conclusion that a simplified, general logic model could be used to depict how WRN (and many NWS activities in general) lead to societal outcomes. This simplified model appears in Figure ES-1. The flow in the model can be described as follows. NWS performs activities that generate outputs. These outputs (warnings, alerts, information from awareness campaigns, etc.) generated by NWS are transmitted to NWS' audiences (partners and the public); these outputs can flow directly to the public or be transmitted (presumably with some value-added) by partners to the public. These outputs should lead to the public being better informed and then to being better prepared for weather-related events. Once a weather-related event is imminent or occurring, the public should then be better able to take appropriate actions which ultimately should lead to reduced injuries and fatalities and reduced property damage.

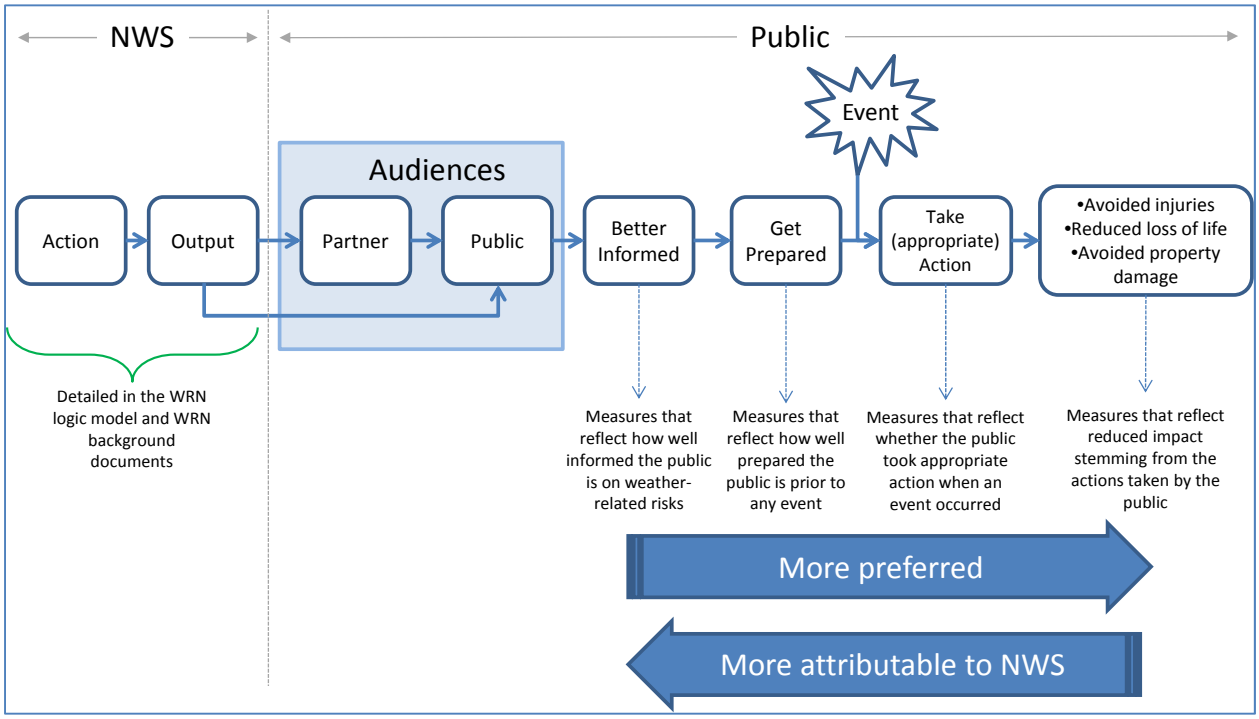

Figure ES-1 – Simplified Logic Model for Generating Societal Outcomes

This simplified logic model led to the development of four categories of metrics to consider:

- Better informed – Measures reflecting how well-informed the public is on weather-related risks.
- Get prepared – Measures reflecting how well prepared the public is prior to any event.
- Take action – Measures reflecting whether or not the public took appropriate actions once an event occurs.

- Avoided property damage, injuries, and fatalities – Measures reflecting the extent to which the actions taken by members of the public reduced property damage or avoided injuries/fatalities.

ERG focused on pilot testing metrics and methods of data collection for each category under this project and, based on the pilot testing, making recommendation for NWS for tracking societal outcomes. It should be emphasized that a key aspect of this project was to pilot test the collection methods and data sources associated with these metrics. Based on discussions with NWS, ERG agreed to focus on four weather event types: winter weather, flash floods, severe thunderstorms, and extreme heat.

The remainder of this executive summary discusses the four metric categories, refines the simplified logic model based on the actual metrics developed, and provides our recommendations for moving forward.

Better Informed

ERG is recommending one metric for this category:

> *Public rating of the extent to which NWS contributes to its understanding of weather dangers*

The data for this metric should be collected and reported on in an event-type specific basis (e.g., for winter weather, flash floods). Data for the metric would come from an NWS website pop-up survey that is currently implemented by NWS and includes a question that would provide data for this metric. The metric would provide a value between 0 and 100 with larger values reflecting better results.

Get Prepared

ERG is recommending one metric in this category:

> *Percent of public that has an emergency preparedness kit*

As with the "better informed" category, the data for this metric should be collected and reported on in an event-type specific basis using the NWS website pop-up survey as the data source. As specified in its definition, the value for this metric reflects a percentage. Larger values of the percentage reflect better results.

Take Action

ERG is recommending two metrics in this category:

> *Percent of public that altered their previous activities after hearing a warning or who remained vigilant*
>
> and
>
> *Percentage of public that sought protective shelter as a first action*

Data for these two metrics should come from implementing the NWS "Quick Response" (QR) surveys for a randomly-selected set of weather events (NWS warnings). ERG has provided details on the random

selection process in the main text of the report. As specified, both reflect a percentage and larger values for both reflect better results.

Avoided Injuries

Our approach to measuring avoided injuries was to compare a statistically-estimated number of "expected" injuries from events to the actual number that occurred. The "expected" number is meant to reflect the amount that could be expected if NWS information did not contribute to people's improved understanding of risk. If the model over-predicts the actual number of injuries (expected exceeds actual), then it can be inferred NWS has contributed to a reduction in injuries. ERG developed estimates for winter weather, flash floods, and severe thunderstorms. Although the modeling was successful, ERG identified a number of concerns with the approach. Thus, we do not recommend collecting data on this category until further work can be done to validate the modeling approach.

Refined Logic Model

Figure ES-2 refines the logic model from Figure ES-1 using the specific metrics discussed above.

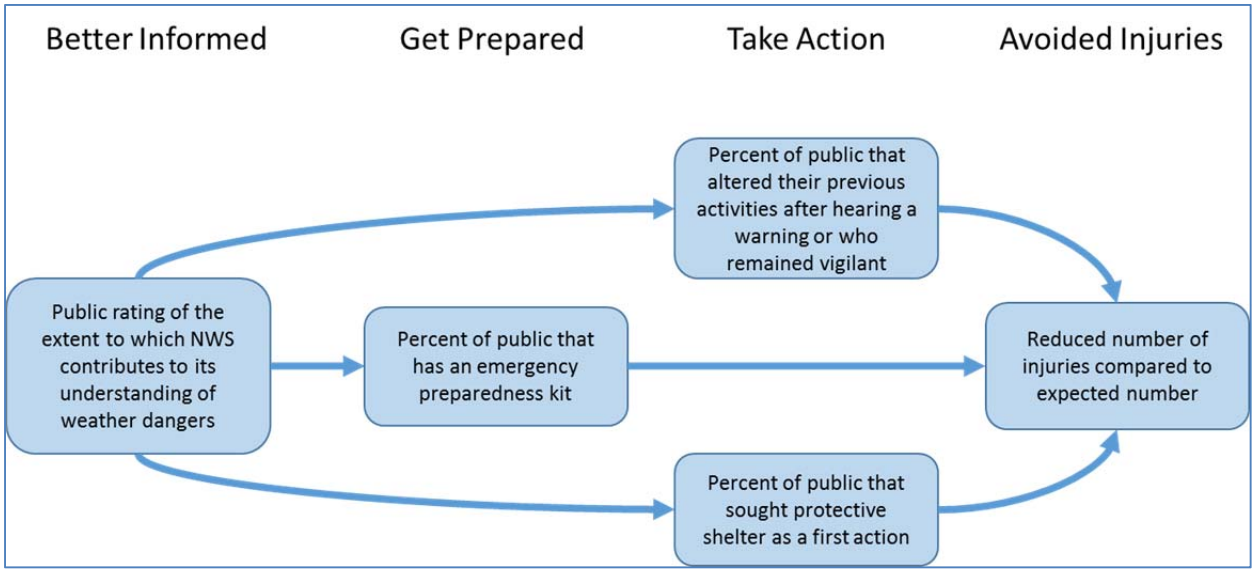

Figure ES-2 – Refined Logic Model Using Specific Proposed Societal Outcome Metrics

Recommendations

This section provides a set of recommendation for NWS to consider in moving forward with developing societal outcome measures.

1. **Collect data on the "better informed," "get prepared," and "take action" metrics, but not the "avoided injuries" metric.** We expect the data we collected for these three metric categories will provide useful information for NWS in tracking societal outcomes. We are not recommending the one metric category that gets closest to a full societal outcome (avoided

injuries). Nevertheless, if the logic model is valid, increases in those who take action should lead to reduced injuries.

2. **Focus on the four weather event types that we have focused on in this report, tracking data for each type separately.** The four events include winter weather, flash floods, severe thunderstorms, and extreme heat. Tracking outcomes from these four event types should provide NWS with an indication of the potential societal outcomes. These four events accounted for 46 percent of injuries and 50 percent of fatalities in 2015 excluding tornadoes from severe weather events; including tornadoes increases those percentages to 89 percent for injuries and 57 percent for fatalities.[1]

3. **For the "better informed" metric, ERG recommends tracking the metric we referred to as "Public rating of the extent to which NWS contributes to its understanding of weather dangers."** To collect data for this metric, NWS should:
    a. Use the question on the extent to which respondents felt NWS contributed to their understanding of the specific weather event.
    b. Collect these data from the website pop-up survey during one quarter of the year.
    c. Translate the values to a metric using the calculation NWS currently uses (see Section 2.2).
    d. Review the demographics for the samples of each quarter to assess whether changes in the metric values are attributable to changes in demographics rather than change in the associated outcome.

4. **For the "get prepared" metric, ERG recommends tracking the metric we referred to as "Percent of public that has an emergency preparedness kit."** To collect these data, NWS should:
    a. Use the question on whether or not the respondents had emergency kits.
    b. Collect these data from the website pop-up survey during one quarter of the year.
    c. Review the demographics for the samples of each quarter to assess whether changes in the metric values are attributable to changes in demographics rather than change in the associated outcome.

5. **For the "take action" metric, ERG recommends tracking two metrics: "Percent of public that altered their previous activities after hearing a warning or who remain vigilant" and "Percentage of public that sought protective shelter as a first action."** To collect these data, NWS should follow the random event (NWS warning) selection process outlined in Section 4.4.2. NWS should also assess whether where the data were collected impacted the values of the metric. For example, collecting these data for winter weather in New England, will result in a set

[1] See Table 14 in the main text for the summarized data derived from http://www.nws.noaa.gov/om/hazstats/sum15.pdf.

of respondents who are familiar with how to deal with winter storms; comparing those results to results from areas less familiar with winter weather events may skew the results.

6. **NWS should perform additional research into ERG's proposed simplified logic model.** Specifically, the logic model we developed is based on our qualitative research into how societal outcomes are generated. The revised version of the model in Figure ES-2 proposes links between the metrics we developed. Further quantitative research is needed to validate the connections in both the original simplified model (Figure ES-1) and the refined model (Figure ES-2).

7. **NWS should perform further research into the avoided injuries modeling approach.** ERG expended as much project resources as possible on this area and we had some success in developing estimates. However, ERG had to limit the resources in this area to allow for collection and analysis of data for other metrics. Additional research is warranted since ERG was able to develop some initial valid models. Further research could help refine and validate the models already developed. One area of potential improvement would be adding in measures of severity for each event type.

1.0 INTRODUCTION

The National Weather Service's (NWS') Weather-Ready Nation (WRN) initiative targets building community resilience in the face of increased vulnerability and risk to weather-related events. The traditional NWS model has relied on providing accurate and timely forecast information to core partners. WRN transforms and extends that model. Accurate and timely information is still essential, but NWS has added requirements to provide interpretation services and to better integrate with its core partners to improve the information provided and to ensure affected communities and individuals have actionable information to inform decisions. One of the key aspects of the WRN initiative is to affect societal outcomes. In other words, it is not enough that the forecast be accurate and timely, the information provided should lead to better decisions that ultimately reduce the impacts of events, including saving the lives of those in harm's way.

Under this project, NWS contracted with Eastern Research Group, Inc. (ERG) to develop a set of potential performance metrics for tracking societal outcomes from the WRN initiative and to collect pilot data for those metrics. This report describes the process we used to develop and pilot test metrics and the resulting data from the pilot testing.

1.1 INITIAL PROJECT WORK AND SIMPLIFIED LOGIC MODEL

ERG's work on this project started with reviewing the NWS logic model and strategy documents, as well as reviewing public literature on the link between weather-related information and societal outcomes. The NWS logic model provided detailed information on how NWS generates forecasts and other information that is disseminated to the public. NWS and ERG agreed that, as a starting point, the project should identify a subset of WRN programs and projects and focus on measuring societal outcomes from those programs and projects. Based on research and discussions with NWS, ERG recommended the following focus areas under WRN:

- The Watch, Warning, Advisory system
- Storm Ready Program
- WRN Ambassadors
- Impact-Based Decision Support Services (IDSS) Pilot Projects
- Awareness Weeks[2]

ERG performed interviews with NWS staff working on each of these areas. Although the interviews we performed were specific to each area, ERG found a number of common underlying themes in how the WRN areas focused on generating societal outcomes. Based on these interviews and discussions with the NWS project staff for this work, ERG developed a simplified logic model that could encompass these common themes; the simplified logic model appears in Figure 1. The flow in the model can be described as follows. NWS performs activities that generate outputs. These outputs (warnings, alerts, information from awareness campaigns, etc.) generated by NWS are transmitted to NWS' audiences (partners and the public); these outputs can flow directly to the public or be transmitted (presumably with some

[2] Now known as "Seasonal Awareness Campaigns."

value-added) by partners to the public. These outputs should lead to the public being better informed and then to being better prepared for weather-related events. Once a weather-related event is imminent or occurring, the public should then be better able to take appropriate actions which ultimately should lead to reduced injuries and fatalities and reduced property damage.

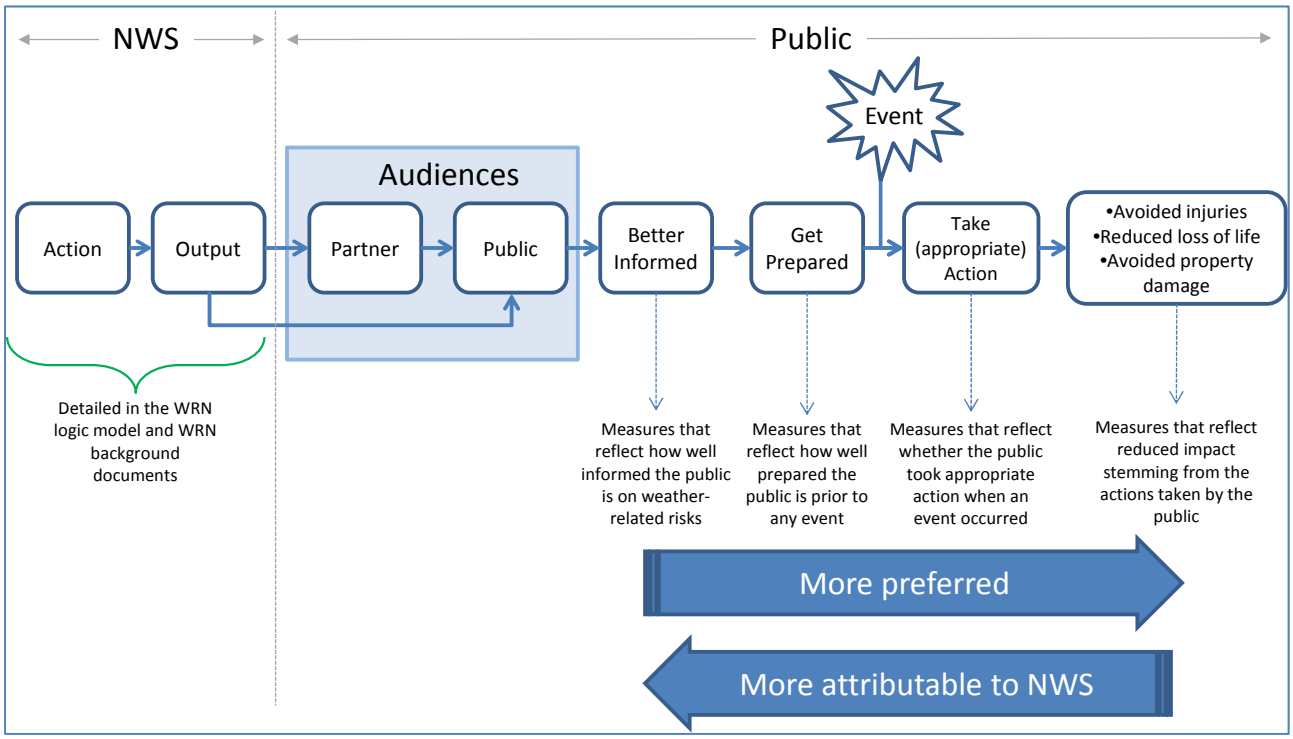

Figure 1 - Simplified Logic Model

1.2 Performance Measure Categories

The original intent of the project was to identify five programs or projects under the WRN program, to develop performance metrics for those programs or projects, and to pilot test a subset of the metrics within the time frame and budget for the project. As we noted, our research into the five programs and projects led us to develop a simplified, more general logic model that applied to all five program and projects. Furthermore, we also expect that the simplified logic model applies to a large number of WRN initiatives, as well as to a large amount of the work NWS performs. That is, in many cases, NWS is providing an output to partners and/or the public and the intent of those outputs are to lead people being better informed, to getting prepared, and to taking protective actions; ultimately leading to reduced injuries, fatalities, and property damage. Thus, we expect that the performance metrics that we developed under this project are applicable to WRN initiative beyond the five that we identified at the start of the project and to many NWS activities.

The simplified logic model assisted ERG in developing a set of four categories of performance metrics based on the flow in the model. These four categories are:

- Better informed – Measures reflecting how well-informed the public is on weather-related risks.

- Get prepared – Measures reflecting how well prepared the public is prior to any event.
- Take appropriate actions – Measures reflecting whether or not the public took appropriate actions once an event occurs.
- Avoided property damage, injuries, and fatalities – Measures reflecting the extent to which the actions taken by members of the public reduced property damage or avoided injuries/fatalities.

Measures reflecting boxes further to the right in Figure 1 are preferred with the ideal measures being those that capture avoided injuries, fatalities, or monetary damages. However, as the boxes flow to the right they become more difficult to attribute to NWS. ERG and NWS agreed to pilot test measures reflecting all four categories.

1.3 Pilot Testing Approach

The remainder of this report describes the pilot testing that ERG performed for the four categories and the conclusions that we drew based on the data that were collected. To ensure the data collection phase was manageable, ERG and NWS agreed, initially, to focus on six weather event types:

- Coastal flooding (including storm surge)
- Flash floods
- Heat and excessive heat
- Hurricane and tropical storm
- Severe thunderstorm and tornado
- Winter storm/winter weather.

The data collection phase of this project was performed in primarily 2015 and early 2016; a time period in which few hurricanes impacted the United States. As such, ERG had few coastal flooding events or almost no hurricanes or tropical storms to use as test cases for the "take action" measures. Additionally, ERG used an NWS survey for the "better informed" and "get prepared" measures and, given the lack of those events, NWS did not focus on either in its own surveys. Thus, coastal flooding and hurricanes/tropical storms were dropped from the in-scope list. Nevertheless, NWS should still consider collecting data for these event types in the future if the number of these types of events warrants collecting the data.

2.0 BETTER INFORMED

The first step in the simplified logic model is that NWS outputs should lead to the public becoming better informed about weather-related risks. This section describes the process for collecting pilot data for this category and defining a metric based on the results of the pilot data collection.

> ***Better Informed*** *– Measures cover how well-informed the public is on weather-related risks*

2.1 DATA SOURCE

NWS has two existing data sources that ERG leveraged for use in this project: a pop-up survey that appears on the NWS website and an associated collection from an online panel. Both surveys contain the same set of questions and focus on one or two types of weather events, rotating the event types out quarterly. Table 1 summarizes the time frame for collecting these data for the four event types we focus on under this project.

Table 1 – Time for NWS to Collect Data from Website Pop-up Survey and Online Panel, by Event Type

Event Type	Calendar Year 2015	Calendar Year 2016
Winter weather	5/9/15 – 7/5/15	1/14/16 – 4/4/16
Flash floods	7/6/15 – 10/6/15	1/14/16 – 4/4/16
Severe thunderstorms	7/6/15 – 10/6/15	-
Extreme heat	10/7/15 – 1/13/16	-

The instruments from these surveys appear in Appendix A of this report. The instruments contain two data elements that are relevant for the "better informed" metric:

- **Knowledge of weather-related events** – "*How would you rate your current knowledge of {EVENT TYPE}-related events on a scale of 1 to 10, where 1 means "very low knowledge" and 10 means "very high knowledge" (i.e., an expert)?*"
- **NWS contributes to understanding of dangers of weather related events** – "*How would you rate the extent to which the information provided by NWS has contributed to your understanding of the dangers of {EVENT TYPE}-related events on a scale of 1 to 10, where 1 means "not at all" and 10 means "significantly"?*"

Data from the first question measures how respondents rate their own knowledge of the different types of weather events. The second question provides an assessment how respondents think NWS has contributed to their understanding of dangers related to the type of weather. Increases in both over time would indicate some improvement in NWS' ability to convey weather information; however, the second question directly ties understanding to NWS activities. On the other hand, the first question can be influenced by non-NWS sources of information. Thus, the second question is consistent with the simplified logic model we developed for this project and we focus on data for the second question in what follows.

2.2 PILOT TESTING APPROACH

To pilot test the "Better Informed" metrics, ERG obtained tabulated results from the online panel and pop-up surveys conducted by NWS for the survey time periods identified in Table 1. One consideration for pilot testing is that the online panel and the pop-up survey are based on different sampling approaches. The online panel is based on selecting a representative sample of the United States whereas the pop-up survey targets individuals who come to the NWS web site. Based on this, it is reasonable to assume that the pop-up survey will be skewed toward individuals who are better informed about weather-related issues than the online panel. Thus, we should expect higher scores for the "Better Informed" category from the pop-up survey compared to the online panel. On the other hand, the online panel will result in a sample of individuals who better represent the U.S. population compared to the pop-up survey. Thus, a measure for being "Better Informed" from the online panel will provide an estimate that is less biased estimate of the how informed the U.S. population is about weather events compared to the estimate from the pop-up survey. The pop-up survey, however, offers a cost advantage over the online panel; obtaining responses from an online panel will result in a cost per respondent. On the other hand, the pop-up survey can be added to the NWS website and responses are collected over time.

2.3 RESULTS AND ASSESSMENT

Table 2 summarizes the data from both the NWS website survey and the online panel by weather event for the question that asked respondents the extent to which NWS contributed to their knowledge of the event. The values in the table reflect a score calculated by CFI Group[3] from the scaled values selected by respondents. For each respondent, the value is calculated as[4]

$$\frac{y-1}{10-1} \times 100$$

where *y* is the scale value (between 1 and 10) selected by the respondent. The value reported in Table 2 is the average across respondents.

The data show a consistency across the three event types in 2015 for which data are available.[5] The same is not true for 2016 where the website survey show appreciably larger scores compared to the online panel.[6] Comparing between 2015 and 2016 for winter weather and flash floods,[7] however, we see declines in three of the four values; there was an increase for the website survey for winter weather. The declines were large for the online panel in both cases; a seven-point decline for winter weather and a six-point decline for flash floods. As noted above, as performance metrics, NWS would be tracking

[3] CFI Group manages these two data collections for NWS.
[4] In words, this is the value selected by the respondent (y) minus the lowest value (1) of scale divided by the highest value of the scale (10) minus the lowest value (1); this essentially normalizes the value selected by the respondent to a percentage of the possible values that could be selected.
[5] These data were not collected for severe thunderstorm events.
[6] ERG does not have access to the raw data and cannot test for the statistical significance of this difference.
[7] These data are currently (at the time of this report writing) being collected for severe thunderstorms and have yet to be implemented for a 2016 survey for extreme heat.

changes in these over time. In the two cases where we have data from two time periods, we are seeing a decline in this score.

Table 2 – Scores for NWS Contribution to Knowledge of Each Event Type, 2015 and 2016 Data Collections

Event Type	2015 Data Collection [a]		2016 Data Collection [a]	
	Website Survey	Online Panel	Website Survey	Online Panel
Winter Weather	80 (n = 3,669)	79 (n = 214)	82 (n = 4,771)	72 (n = 246)
Flash Floods	76 (n = 1,420)	76 (n = 152)	75 (n = 6,204)	70 (n = 487)
Severe Thunderstorms	[b]	[b]	[c]	[c]
Extreme Heat	63 (n = 4,771)	65 (n = 246)	[c]	[c]

[a] Number of respondents appear in parentheses below each score value.
[b] Not collected during this implementation.
[c] Not applicable.

This analysis raises a concern about the comparability of the online panel data between 2015 and 2016. Specifically, the website survey shows a consistency in the values between the two years, but the online panel shows a large decline between 2015 and 2016. In short, ERG has a concern over why the online panel would show an appreciable decline, but the website survey would not. To avoid this, ERG recommends that in future implementations of the online panel, NWS request the selected panel to match previously used panels in terms of demographics as much as possible.

2.4 RECOMMENDATIONS

Table 3 provides a proposed metric for "better informed." ERG's recommendation is to formulate a metric called "Public rating of the extent to which NWS contributes to its understanding of weather dangers." This would be calculated from the website survey only using the question identified above. Given that the questions being asked are specific to event types, NWS should report on these metrics based on event type. For example, two metrics to report on would be:

- Public rating of the extent to which NWS contributes to its understanding of **winter weather** dangers
- Public rating of the extent to which NWS contributes to its understanding of **flash floods** dangers

Table 3 – Proposed Metrics for "Better Informed" Category

Metric	Collection Approach	Question in Survey	Calculation
Public rating of the extent to which NWS contributes to its understanding of weather dangers	NWS website survey, collected once per year over a three-month period	How would you rate the extent to which the information provided by NWS has contributed to your understanding of the dangers of {EVENT TYPE}-related events on a scale of 1 to 10, where 1 means "not at all" and 10 means "significantly"?	Average score from survey where the score is calculated for each respondent as the respondent's rating minus one divided by 9 and then multiplied by 100.

Furthermore, we are recommending that the data come from the web survey and not the online panel. This reflects that the idea that the website survey may be economical long term. The website survey rotates questions quarterly and NWS should continue to rotate in questions for the four weather events we have focused on in this report: winter weather, flash floods, extreme heat, and severe thunderstorms. It is possible that the demographics of those who take the website survey will shift over time, resulting in changes in the metric that are based on demographic-based changes. To assess this, NWS should review the demographics from the website survey over time. Finally, we also recommend using with the calculation approach employed by CFI Group and provided in equation form above.

3.0 GET PREPARED

The "get prepared" measures flow from the "better informed" measures on the simplified logic model. That is, people who are better informed should be more likely to take proper actions to be prepared for weather events.

> ***Get Prepared*** *– Measures that reflect how well-prepared the public is prior to any event*

3.1 DATA SOURCE

Once again, ERG used data from the NWS online panel and website pop-up survey discussed in Section 2.1 above; the survey instruments for these can be found in Appendix A. ERG considered three questions from the surveys as potential sources for this metric:

- Whether or not the respondent had a hazardous weather safety plan
- Whether or not the respondent had an emergency preparedness kit
- The items that respondents had in their emergency preparedness kit[8]

To determine which of these to recommend, ERG reviewed data from the initial data collections covering winter weather and flash floods performed by NWS in early 2015. There were some slight differences in how the winter weather-focused and flash flood-focused surveys were implemented in those data collections. First, for winter weather, the presence of an emergency preparedness kit was only asked if the respondents indicated they had a plan. NWS altered this and asked about the existence of a kit from all respondents in the flash flood-focused survey based on a recommendation by ERG. Second, to stay within question number limits,[9] the flash flood survey eliminated the question about the emergency preparedness plan.

Our initial analysis of those two data collections appeared in our September 8, 2015 piloting results summary memo. Based on the data at that point, ERG recommended focusing on whether or not the respondents have emergency kits. First, a "safety plan" can have multiple interpretations; for example, a plan could mean a specific set of things to do in case of emergency or it could mean a general idea of what do in case of emergency. The items in the emergency preparedness kit question was also problematic since it was difficult to form into a coherent metric. ERG used the number of items in the kit, but the number of items ignores the importance of different items. Thus, ERG determined that whether or not a respondent had an emergency preparedness kit as being the best approach for the "get prepared" measure. For one, in contrast to a plan, a "kit" usually refers to a specific set of items.

[8] For winter weather, the list included: windshield scraper and brush or small broom; sack of sand or cat litter for traction; shovel; tow rope or chain; battery booster cables; mobile phone charger; blankets or sleeping bag; flashlight; first aid kit; water; high-calorie, non-perishable food; extra clothing to keep dry. For flash floods, the list included: water; food; a battery-powered or hand crank radio; flashlight and extra batteries; first aid kit; whistle to signal for help; moist towelettes, garbage bags and plastic ties for personal sanitation; wrench or pliers to turn off utilities; manual can opener for food; local map; and cell phone with chargers. Both lists were based on information from the Federal Emergency Management Agency (FEMA).

[9] The survey is question-limited due to cost concerns. NWS needed to add in other questions to the survey and since the emergency preparedness kit question was being expanded to all respondents, NWS decided to eliminate the safety plan question.

Second, those who have a kit can be seen as more prepared than those who do not have one, regardless of the items in the kit.

3.2 Pilot Testing Approach

As with the "Better Informed" metric described above, ERG obtained tabulated data from NWS for the relevant questions. Since this metric also uses the online panel data and the pop-up survey data, the same caveats as discussed in Section 2.2. are applicable here. That is, we might reasonably expect the respondents to the pop-up survey to be more prepared compared to those for the online panel. On the other hand, the respondents to the online panel should provide a less biased estimate of the level of preparedness compared to the pop-up survey. Finally, we also expect the pop-up survey provides a more cost-effective means of collecting data.

3.3 Results and Assessment

Table 4 summarizes the data collected on the percentage of respondents who had emergency kits for winter weather and flash floods from the two surveys time frames that covered these event types. These same data have not been collected for extreme heat or severe thunderstorms. Table 4 also provides results from statistical tests comparing the difference between the website and online panel surveys and from comparing the two survey time frames.[10]

The statistical tests show that there are statistically significant differences between the website survey and the online panels for both winter weather and flash floods. For winter weather, the website survey respondents had statistically significant larger proportions with emergency kits for both survey time frames. For the flash floods, however, the reverse was true and the online panel respondents had statistically significant higher proportions with kits compared to the website survey. The results for flash floods runs counter to our expectation (see Section 3.2) that the website survey respondents would have higher levels of preparedness.

Furthermore, given that metrics are intended to be tracked over time, we can assess temporal changes. Here we see no statistically significant difference in three of the four cases; the one exception being flash floods for the website survey. The flash flood data also exhibit large declines in the percentages: an eight percentage point decline for the website survey between the two years (statistically significant) and a 10-point decline for the online panel between the two years (not statistically significant). Thus, for flash floods, even though only one decline was statistically significant, there does appear to be a decline in preparedness over time as evidenced by both.[11]

[10] In contrast to the "better informed" metric, ERG can perform these statistical tests since we do not need access to the raw data to perform statistical tests for percentages; standard errors can be derived from the mean percentage values.

[11] It is possible, however, that the change here could be related to changes in the demographics of the samples.

Table 4 – Summary of Results from NWS Website Survey and Online Panel for Whether or Not Respondents Have an Emergency Preparedness Kit

Weather Event Type and Survey Time Frame	Website Survey	Online Panel Survey	Significant Difference Between Website and Panel Surveys? [b]
Winter Weather			
May – July 2015 Survey	57% [a] (n = 3,132)	41% [a] (n = 120)	Yes (3.47)
January – April 2016 Survey	59% (n = 3,669)	44% (n = 214)	Yes (4.32)
Statistical difference between survey time frames?	No (1.67)	No (0.53)	-
Flash Floods			
July – October 2015 Survey	31% (n = 3,743)	41% (n = 101)	Yes (2.14)
January – April 2016 Survey	23% (n = 1,420)	31% (n = 152)	Yes (2.20)
Statistical difference between survey time frames?	Yes (5.67)	No (1.63)	-

[a] These were not the values reported from the survey. In the survey, only respondents who said they had a winter weather safety plan were asked if they had an emergency preparedness kit. We adjusted the percentage to reflect this by multiplying the percentage reported in the survey for this question by the percentage that have a safety plan; this adjustment essentially scales the percentage back to whole sample. However, there may be some respondents who do not have a plan, but have a kit; our adjustment assumes that those without a plan also do not have a kit.
[b] The number in parentheses is the z-score for a test of a statistical difference between two proportions at a 95 percent level of confidence.

3.4 RECOMMENDATIONS

Table 5 provides the proposed metrics for the "get prepared" category. ERG's recommendation is to focus on the percentage of the public that has an emergency preparedness kit. A "plan" can have many interpretations by respondents. A plan can be comprehensive or simple; it can be well thought out or can be based on what the respondent interprets as "common sense." However, a kit is more concrete; it is a collection of items that can be used in an emergency.

Table 5 – Proposed Metrics for "Get Prepared" Category

Metric	Collection Approach	Question in Survey	Calculation
Percent of public that has an emergency preparedness kit	NWS website survey, collected once per year over a three-month period	Do you have a {EVENT TYPE} emergency preparedness kit for your vehicle? (Yes/No)".	Percentage that answer yes to the question.

As with the "better informed" metric category, ERG recommends that this be tracked in an event type-specific manner. For example, the trackable metrics from the data collected by NWS so far are:

- Percent of public that has a **winter weather** emergency preparedness kit
- Percent of public that has a **flash flood** emergency preparedness kit

ERG recommends that these data be collected annually using the website survey as the means of collecting these data. The NWS website survey rotates questions quarterly and NWS should continue to rotate in the question on emergency kits for the weather events we have focused on in this report.

Additionally, although the online panel is designed to collect representative data from the U.S. population, we expect the website data will prove to be more cost-effective over the long term. Even though the website data may be skewed to people who are more weather-savvy since the survey targets NWS website users, as long as the data are collected consistently over time, the data will be comparable over time. Thus, the data may not provide an accurate estimate of the percentage of the U.S. population that has emergency kits, but will provide useful data for tracking performance over time. As with the "Better Informed" category, we also recommend assessing the sample over time for changes in demographics that could influence changes in the metric value itself.

4.0 TAKE ACTION

Take Action *– Measures that reflect whether or not the public took appropriate action once an event occurs*

This section describes the approach to collecting data for the take action measure in the simplified logic model; individuals who get prepared should be more likely (or more able) to take appropriate actions once an event occurs.

4.1 DATA SOURCE

ERG collected data for this this category using the NWS "Quick Response Surveys" (QR surveys). The QR surveys provided a unique opportunity to collect data under this project since (1) NWS had already spent time and effort into developing the questions for the survey instruments and (2) they were covered by an existing OMB approval. The full survey instruments for the events we included in our analysis appear in Appendix B of this report. The QR surveys contained four data elements (in bold in the bullets that follow) that we expect can contribute to a "take action" metric:

- *Did you take any of the following actions when you received the {EVENT TYPE} warning?* This question contained several options for respondents and respondents could select "yes," "no," or "I don't remember" for each one. Two of the sub-questions are of interest:
    - **Sought shelter** – Each event type survey has a question on whether the respondent sought shelter. For the extreme heat survey, we used the "remained indoors" question as the seeking shelter question; we interpret a "yes" to this question as the respondent taking a protective action. For winter weather, we used the percentage that indicated they altered their routine as the seeking shelter question.
    - **Continued to do previous activities** – Each survey instrument also asks respondents if they continued to perform their previous activities after hearing the warning. We interpret a "yes" response as the respondent indicating that the warning did not alter their normal routine, possibly an indication they did not take the warning seriously enough.
- *Which of the following actions was the first action you took when you received the {EVENT TYPE} warning?* This question had several options and the respondent could only choose one. We see two response options from this question as most relevant:
    - **Sought shelter (first action)** – If the respondent's first action was to seek shelter, we interpret this as the respondent taking a protective action as his/her first step in response to the warning.
    - **Did nothing (first action)** – If the respondent's first action was to "do nothing," we interpret this as a sign that the respondent took no protective action when the warning was issued.

For the "sought shelter" data elements, a positive result would be indicated by larger percentages; as noted in the bullet points above the question was asked differently for extreme heat and winter weather compared to the other events we looked at. For the "continued to do previous activities" and "did nothing as first action," a positive result would be indicated by lower percentages.

4.2 Pilot Testing Approach

Working with NWS, ERG selected four warnings issued by NWS in June to August of 2015 and a fifth one in January 2016; these included a flash flood warning, a severe thunderstorm warning, a tornado warning, an extreme heat warning, and a winter weather warning. [12] ERG used the list of counties in each warning product to develop a geographic area to use in collecting the data. ERG provided the list of counties to Qualtrics, Inc. who obtained an online panel to use for each data collection. ERG coded the survey in our Qualtrics account and Qualtrics implemented the survey with the online sample for each. Table 6 summarizes information on the five surveys that were implemented for this metric category.

Table 6 - Surveys Implemented by ERG for the "Take Action" Metric

WFO	Product (Warning Type)	Warning Issuance Date	Survey Dates	Number of Respondents
Slidell, LA	Flash Flood	6/9/15	6/29/15 – 7/1/15	125
Jacksonville, FL	Severe Thunderstorm	6/9/15 – 6/10/15	6/30/15 – 7/1/15	127
Taunton, MA	Tornado	6/23/15	7/1/15 – 7/5/15	128
Phoenix, AZ	Extreme Heat	8/15/15	8/24/15 – 8/27/15	160
Sterling, VA	Winter Warning	1/21/16 – 1/22/16	2/1/16 – 2/2/16	180

4.3 Results and Assessment

Table 7 summarizes the results for the four data elements identified above from the five surveys that we implemented. There is variation in the "sought shelter" element with only 26 percent seeking shelter in response to the flash flood survey, but 78 percent "remaining indoors" for the extreme heat survey. "Remaining indoors," however, poses a different risk for those who do not have air conditioning for a heat event.[13] We do not recommend summing the seeking shelter element across event types. Similarly, the seeking shelter as a first action also has variation across the event types that is most likely due to how people respond to the specific events.

The percentages that continued previous activities and that did nothing as a first action, on the other hand, have more stable percentages across the event types. These two data elements, however, tell different stories. The "continued previous activities" element shows high percentages. In fact, several people in each survey said they continued previous activities and also selected one or more protective measures from the list. The "did nothing as a first action" element, on the other hand shows very low percentages.

[12] The time lag between the first four and the fifth one reflect ERG and NWS waiting for a winter weather warning that was substantial enough to warrant a survey. The event selected was the January 21-22 snow event that impacted the Washington, DC area commonly referred to as "snowzilla."

[13] In response, NWS is changing this action to "seeking air conditioning" in the next version of the survey that includes heat as an event.

Table 7 – Summary of Results from Quick Response Surveys for Data Related to "Take Action" Metric

Data Element	Slidell, LA – Flash Flood	Jacksonville, FL – Severe Thunderstorm	Taunton, MA – Tornado	Phoenix, AZ – Extreme Heat	Sterling, VA – Winter Weather
Percentage that sought shelter	26% (n=120)	53% (n=117)	39% (n=126)	78% [a] (n=148)	65% [b] (n = 174)
Percentage that continued previous activities	71% (n=120)	65% (n=116)	60% (n=124)	59% (n=147)	49% (n=172)
Percentage that sought shelter as first action	10% (n=125)	25% (n=127)	13% (n=128)	40% [a] (n=160)	24% [b] (n=177)
Percentage that did nothing as first action	14% (n=125)	11% (n=127)	13% (n=128)	15% (n=160)	9% (n=177)

[a] For extreme heat, the questionnaire asked respondents if they remained indoors.

[b] For winter weather, the questionnaire asked respondents if they altered their plans.

4.4 Recommendations

4.4.1 Metrics

Table 8 provides our proposed metrics for the "Take Action" category. We are recommending two metrics. The first metric we recommend is the percentage of people that alter their activities as a result of a warning or who (at least) remain vigilant about potential weather hazards.[14] This will need to be measured from two survey questions. The first part of the metric would be calculated as the percentage that say "no" to "continuing previous activities." The second part would require adding a question to the survey that asks if the respondent started or continued to monitor weather forecasts or other sources of information (e.g., social media, friends, family, etc.). Thus, this metric would then be defined as any respondent who (a) answered "no" to continuing previous activities and (b) answered "yes" to starting or continuing to monitor weather forecasts.

The second metric we recommend is the percentage of people who seek shelter as their first action. As with the previous metrics, we recommend these be calculated for the event types separately.

[14] The second part of this metric is based on the idea that individuals can continue their previous activities, but remain vigilant or concerned about hazardous weather events. For example, someone at an outdoor event that hears a severe thunderstorm warning can continue their participating in the event, but also begin to monitor weather forecasts.

Table 8 - Proposed Metrics for "Take Action" Category

Metric	Collection Approach	Question in Survey	Calculation
Percent of public that altered their previous activities after hearing a warning or who remained vigilant	Survey – online panels implemented for a random selection of warnings issued	*Did you take any of the following actions when you received the {EVENT TYPE} (advisory/warning)? Did you continue previous activities? Did you monitor weather forecasts? [a]*	Percentage that answer **no** to the question on continuing previous activities or who answer **yes** to the question on monitoring weather forecasts.
Percent of public that sought protective shelter as a first action	Survey – online panels implemented for a random selection of warnings issued	*Which of the following actions was the first action you took when you received the {EVENT TYPE} warning? Response option: seek shelter [b]*	Percentage that select "seek shelter" (or option determined to be seeking shelter).

[a] The second question about monitoring weather forecasts is not currently part of the questionnaire; ERG is recommending including this question in future surveys.
[b] For extreme heat, the wording NWS should be "seek air conditioning." For winter weather, the wording should be "altered your routine."

4.4.2 Data Collection

These two metrics will require collecting data from a set of randomly selected warnings (events) that occur using the QR surveys that we used in this project. In this section, ERG provides a set of steps to use in selecting these events. In our pilot data collections, we collected between 125 and 180 responses for the five warnings that we selected. For a full implementation, ERG suggests collecting at least 328 responses annually for each event type spread over eight warnings issued by WFOs (approximately 42 per event) for winter weather, severe thunderstorms, and flash floods and over six warnings (approximately 55 per event) for extreme heat events. The surveys to collect the data should be implemented within one to two weeks following the events that are selected. The 328 responses will allow for testing for a statistically significant difference between two years at 90 percent confidence assuming a five percentage point increase in the metric. That is, if the metric increased by five percentage points (or more) between two years, a sample of 328 respondents each year would discern that change as a statistically significant increase at the 90 percent level of confidence.[15] NWS should also assess whether the changes in the metrics are associated with where the data were collected; for example, collecting data on winter weather from New England would result in a set of respondents who are familiar with how to respond to winter weather events.

- **Step 1 – Select WFOs**. At the start of the fiscal year, randomly select four WFOs for winter weather, severe thunderstorms, and flash floods and three WFOs for extreme heat to include in the data collection process. For each event type, only WFOs that experience the event type frequently should be included in the random selection process.[16] ERG suggests alerting the WFO of its inclusion on the process. Random selection should be done using a "probability proportional to size" approach; that is, the more events that a WFO experiences, the more likely

[15] A 328 respondent sample would also allow for detecting a 6.4 percentage point increase as statistically significant at the 95 percent level of confidence.
[16] ERG recommends that NWS determine the criteria for a WFO to be included in the random selection process; however, we recommend that only WFOs who are very likely to experience more than five events be included in the data collection process. The reason for having at least five events will be made clear in the third step.

it should be included in the sample.[17] Furthermore, the sequence in which the WFOs were selected should be tracked since this will be used in Step 2.

- **Step 2 – Select months**. The data collection process should be spread out over the months in which the event type is most likely to occur. ERG reviewed NWS data on warnings from the Interactive Product Database and has made recommendations on which months to include for each event type and how to assign WFOs to each month; these recommendations are:
    - Flash floods: May – August. Assign the WFO that was selected first in Step 1 to May, the second one selected to June, etc.
    - Severe thunderstorms: May – August. Assign the WFO that was selected first in Step 1 to May, the second one selected to June, etc.
    - Winter weather: December – February. Assign the first selected WFO to December, the second two selected to January, and the final one selected to February.
    - Extreme heat: July and August. Assign the first selected WFOs to July and the second two selected WFOs to August.
- **Step 3 – Select events**. The first two "in-scope relevant" events that occur for each selected WFO should be the events selected for the survey.[18] NWS should use its best judgement on what to classify as an "in-scope relevant" event for these surveys.[19] If the WFO has less than two events in the month, events that occur in the next month are fine to select also.
- **Step 4 – Implement the QR surveys**. NWS should implement the appropriate QR surveys using an online panel provider such as Qualtrics or the one used by CFI Group. The surveys would be implemented using the same approach ERG used above; specifically, once a warning is issued, the survey will be implemented for the counties covered by the survey.[20]

[17] For example, for simplicity, assumes there are three WFOs to select from and WFO #1 experiences 100 events per year, WFO #2 experiences 75 events per year, and WFO #3 experiences 25 events per year. In this case, WFO #1 should have a 50 percent probability of being selected (100 events at WFO #1/200 total events), WFO #2 should have a 37.5 percent probability of being selected (75/200), and WFO #3 a 12.5 percent probability of being selected (25/200).

[18] Deviations from selecting the first two events are fine also. For example, if NWS does not survey for the first event to occur at a WFO in specific month due to time constraints, NWS should just select the next two that occur. It is important, however, to avoid non-random selection criteria; for example, selecting an event because the WFO felt it "did a good job" for a particular event would bias the results.

[19] As an example, as ERG was waiting to identify a winter weather event to survey under this project, we proposed two potential events to survey prior to the one finally used. NWS rejected the first two we proposed because each focused on winter weather warnings for higher elevations and would have not impacted many people. Additionally, when ERG selected the Baton Rouge flash flood event, NWS staff reviewed the area carefully to ensure the area affected was not a low population area.

[20] NWS can also refine the targeting of these surveys by including additional geographic detail if available. For example, if NWS has information on the zip codes within the county that are affected, then zip codes can be used to target the survey.

5.0 AVOIDED INJURIES

Avoided injuries *– Reflects the extent to which actions taken by public led to avoided injuries*

5.1 APPROACH

Our approach to piloting a data collection for the avoided injuries category took a different approach compared to the other three categories. Instead of collecting primary data for this category, ERG used data that NWS currently collects on weather events (Storm Data storm events database) and combined those data with Census Bureau data to perform statistical analyses. For this category, we developed models for severe thunderstorms, flash floods, and winter weather as part of the pilot data collection.[21]

To develop a metric for the number of *avoided* injuries, we need to develop an estimate of the expected number of injuries to compare to the actual number that occurs. The "expected number" of injuries should reflect the situation where no improvements are made; that is, the expected number should reflect "if NWS has made no improvements to the information it provides, then we should expect *x* number of injuries in {year}." If we have the actual number of injuries in "{year}", then we can compare the expected number to the actual number of injuries to determine whether injuries from weather events are lower than expected. There are, naturally, a number of details that go into this calculation:

- Different weather events need to be treated separately. For example, winter weather-related injuries should be treated separately from severe thunderstorm injuries. In our work we have separated events into the following categories:[22]
    - Flash floods
    - Heat and extreme heat
    - Thunderstorms/severe weather
    - Winter weather
- We need a method to calculate the number of expected injuries. This is discussed in detail below and has been the focus of ERG's statistical modeling efforts under this metric.

We calculate avoided injuries using the following approach:

1. Develop a statistical model for a baseline period that relates the number of injuries that occur during weather events to factors we would expect to contribute to injuries. The statistical model needs to account for the fact that only a small number of events actually have a recorded number of injuries. We also removed outlier observations from the analysis at this point; we defined outliers as any event that had a strong influence on the results.[23, 24]

[21] The statistical models for extreme heat were not valid once estimated.
[22] As mentioned in footnote 21, however, the statistical models for extreme heat were not valid.
[23] For example, for flash floods we removed the three events with the largest number of injuries: 24, 30, and 136 injuries, respectively. Including these three events resulted in significantly inflated estimates of the number of avoided injuries.

2. Use the baseline period model to predict the number of injuries for events occurring after the baseline period during a "measurement period."
3. Compare the predicted number of injuries to the actual number for the measurement period. If the predictions using the baseline model over-predict the number of injuries that actually occur, then the state of the world has improved; that is, fewer injuries occurred than we would have expected to occur given the baseline model. On the other hand, if the model predicts fewer injuries than actually occur, the state of the world has not improved; that is, the model is telling us we expected fewer injuries than actually occurred.

For this analysis, we set the baseline period equal to 2007-2011 and we used 2012 and 2013 as the measurement periods.

5.2 Data Source

ERG downloaded data from the Storm Data storm events database for the weather events defined above. The data we downloaded contained records for verified events from 2007 to 2014. Each event in the database has information on its location (e.g., state, county, etc.). We merged these data with Census Bureau data for the following information:

- County population
- County land area (used with population to form a measure of population density)
- County median income
- Average age of housing in the county

A key determinant of injuries from a weather event should be the severity of the event; however, these data are not easily measured for some event types. For severe weather (thunderstorm winds), the Storm Data storm events database tracks wind speed for the event and we use wind speed as a measure of severity. We do not, however, have measures of severity for flash floods or winter weather in our data sources.

5.3 Approach: Estimating a Baseline Statistical Model

ERG estimated a baseline statistical model that related the number of injuries in each weather event to the following factors:

- Total population in the county (measured as a natural log) – The more people that live in a county, the more people there are to be harmed by any given event.
- Population density in the county – The more "packed" a county is, the more likely that people will be harmed.[25]

[24] Should NWS implement this method for calculating avoided injuries in the future, NWS will need input from a statistician on implementing the modeling approach and in determining if outliers exist in a specific analysis and whether to remove those outliers.

[25] Total population and population density actually have only a modest correlation in the data that we used (correlation coefficient < 0.4); thus we are able to include both in the same statistical model.

- County income (measure as a natural log) – More affluent counties may have better access to protective measures.
- Average age of the housing – Counties with, on average, older homes may be more likely to see injuries.
- For severe thunderstorms, magnitude (wind speed) – The more intense the winds in a thunderstorm, the more likely there are injuries.
- Control variables reflecting specific years – We formulated a set of binary variables for each year to control for trends from year to year. For example, we have a variable called "Y2008" that is set equal to one for any event in 2008 and zero for events in other years. In estimating the statistical model, we used controls for 2007-2010, effectively setting 2011 as the "base case" in the model.
- Control variable reflecting the NWS region – We formulated a set of controls for the NWS regions. In the models, we used the controls for the Central Region, Southeast Region, and the Eastern Region; making all other areas the base case for the model.

Some of our preliminary statistical estimates also included a time trend (instead of the binary year controls), but these analyses did not improve on the model described in the bullet points. For the winter weather analyses, we added in binary control variables for different types of winter weather (blizzards, heavy snow, and winter storm).

The statistical method is a complex statistical model referred to as a "zero-inflated Poisson" (ZIP) regression model. A Poisson regression model is used in cases where the outcome variable (injuries in our case) reflects a count of items; that is, the outcome is integer values, usually smaller values, and non-negative. Poisson models, however, need to be adjusted if the number of zero values in the outcome variable is large; the adjustment is referred to as "zero inflation." In short, a ZIP model treats the zero values and the non-zero values as coming from different statistical distributions and estimates two separate models. An analogy may make this clearer; in studying consumer purchases, people first make a decision on whether to purchase something (i.e., a zero or non-zero amount) and, if the they decide to purchase, then decide on how much to purchase (the value of the non-zero amount). In the case of weather events, there are factors that lead there to be either no injuries reported or some positive number of injuries reported during an event and, when there are injuries reported, there are factors that lead to more injuries being reported.

ERG estimated ZIP models for injuries[26] occurring during thunderstorms, flash floods, and winter weather events using the factors in the bullet points above as explanatory variables; separate models were estimated for each event type separately. The models were estimated over 2007 to 2011 as a baseline period using 33,074 thunderstorm wind events (345 with non-zero numbers of injuries), 7,589 flash food events (38 with non-zero number of injuries), and 7,226 winter weather[27] events (142 with

[26] The Storm Data storm events database data provides counts of "direct" and "indirect" injuries from weather events; we used the combined sum of direct plus indirect in our analyses.
[27] This included events categorized as "winter weather," "winter storm," heavy snow," and "blizzard."

non-zero number of injuries). Model diagnostics indicated all of the models were valid.[28] It should be noted that there were 80,651 total thunderstorm wind events, 29,081 total flash flood events, and 53,338 total winter weather events over that time period. The events that were excluded from the model were excluded due to missing data for at least one explanatory factor (see bulleted list above).

5.4 RESULTS

This section summarizes the results for severe thunderstorms, flash floods, and winter weather.

5.4.1 Severe Thunderstorms

ERG used the estimated baseline model for severe thunderstorms to predict the number of injuries for events occurring in 2012 and 2013.[29] We then compared this to the number of injuries for thunderstorm wind events in 2012 and 2013; the results appear in Table 9 below. As can be seen in the table, the model predicts fewer injuries in 2012 than actually occurred and then predicts more injuries in 2013 than actually occurred. Overall, this approach estimates that about 63 fewer injuries occurred in 2012-2013 than expected.

Table 9 – Estimated Avoided Injuries for Severe Thunderstorm Events, 2012-2013

Year	Number of Thunderstorm Wind Events Used in Analysis [a]	Actual Number of Injuries for Thunderstorm Wind Events [a]	Predicted Number of Injuries for Thunderstorm Wind Events Using Baseline Model	Estimated Avoided Injuries
2012	7,072	204	180.3	-23.7
2013	6,100	72	159.1	87.1
Total	**13,172**	**276**	**339.4**	**63.4**

[a] ERG used only events where a predicted value could also be calculated.

5.4.2 Flash Floods

The estimates for flash floods appear in Table 10. ERG developed two sets of estimates. First, we estimated the model using all events with complete data for the variables used in the model. Second, we excluded three events we expect are outliers. For the estimates that included the outliers, the modeling approach predicts 15 avoided injuries. For the estimated that excluded the outliers the modeling approach predicts about 7 fewer injuries.

The outlier events were excluded since they exerted a strong influence on the final results; as can be seen in Table 10, including these three events more than doubled the estimated number of avoided injuries. These three events involved 24, 36, and 136 injuries, respectively, and represent the top three

[28] The severe thunderstorm model had a Wald Chi-squared value of 31.73 which was significant at the one percent level. The flash flood model had a Wald Chi-squared value of 44.7, significant at the one percent level. The winter weather model had a Wald Chi-squared value of 486.7, significant at the one percent level.

[29] There was one complication in making these predictions: the baseline model uses binary control variables for years and we need to predict for years outside of the range of years we are using in the baseline. We overcame this by using the 2011 as the base case in the model. This essentially assumes that 2012 and 2013 will be similar in injury trends to 2011.

events in terms of injuries in the data. Once we removed these outlier events, the model stabilized and removal of additional higher-injury events resulted in little change to the final estimates.

Table 10 – Estimated Avoided Injuries for Flash Flood Events, 2012-2013

Year	Number of Flash Flood Events Used in Analysis [a]	Actual Number of Injuries for Flash Flood Events [a]	Predicted Number of Injuries for Flash Flood Events Using Baseline Model	Estimated Avoided Injuries
Baseline model using flash flood events with complete data [b]				
2012	1,281	8	16.6	8.6
2013	2,170	12	18.7	6.7
Total	**3,451**	**20**	**35.4**	**15.4**
Baseline model using flash flood events with complete data, but excluding outlier events [c]				
2012	1,281	8	11.6	3.6
2013	2,170	12	15.1	3.1
Total	**3,451**	**20**	**26.8**	**6.8**

[a] ERG used only events where a predicted value could also be calculated.
[b] "Events with complete data" refers to an event that had non-missing data for the variables used in the analysis; see bulleted list in Section 5.3.
[c] ERG excluded three events due to their overly-strong influence on the final results. These three events had reported injuries of 24, 30, and 136 injuries, respectively.

5.4.3 Winter Weather

The winter weather-related estimates appear in Table 11. ERG estimated that approximately 56 fewer injuries occurred in 2012-2013 than would have been expected based on the baseline model.

Table 11 – Estimated Avoided Injuries for Winter Weather Events, 2012-2013

Year	Number of Winter Weather Events Used in Analysis [a]	Actual Number of Injuries for Winter Weather Events [a]	Predicted Number of Injuries for Winter Weather Events Using Baseline Model	Estimated Avoided Injuries
2012	912	47	58.2	11.2
2013	1,372	52	96.9	44.9
Total	**2,284**	**99**	**155.1**	**56.1**

[a] ERG used only events where a predicted value could also be calculated.

5.5 Assessment

Overall, the results provide some positive results in that we were able to successfully estimate a positive number of avoided injuries for each event we considered. There are, however, a number of reasons to be concerned about this construction of an avoided injuries measure, including:

- **The estimates are based on a complex statistical design** – We used a zero-inflated Poisson (ZIP) model to develop our estimates. Although the ZIP model is an accepted statistical practice, the method is still a complex procedure. The use of a complex method to develop a performance metric may reduce its acceptability.

- **The analysis is based on several weather events where zero injuries are reported** – The number of zero-event records in the data means that the estimated models rely on few events (relative to the total number) to discern patterns in the data when injuries occur. The use of a ZIP model should account for some of this, but it should still cause some concern.
- **The data are based on numbers of injuries reported by WFO** – There may be some events where injuries occur, but are not reported.
- **Several events are excluded due to data limitations** – As noted above, the statistical modeling had to exclude large numbers of events due to missing data for some events; that is, the county was not in the Census Bureau data sources for some data elements we used.
- **No measure of severity** – We have no real measure of severity for flash floods or for winter weather.
- **No guarantees that the estimated number of avoided injuries will be positive** – The baseline model could also predict fewer injuries than actually occur as was the case for severe thunderstorms in 2012.
- **Counterintuitive statistical estimates in some cases** – The statistical analyses produced counterintuitive results in some cases. For example, in our winter weather model results, population was found to be significantly negatively associated with injuries.
- **Will need continual updating** – Using this type of measure will require updating the baseline analysis every 2-3 years.

Thus, our overall assessment of this category of measures, and our approach, is that further assessment is needed prior to using this as a measure.

6.0 SUMMARY AND RECOMMENDATIONS

6.1 SUMMARY

Table 12 summarizes the proposed metrics following the pilot testing on this project. Figure 2 provides a logic model that includes the metrics within each category. This updated logic model modifies the model in Figure 1 by disconnecting the linkage between "get prepared" and "take action" since we do not think our "get prepared" metric would contribute directly to the "take action" metrics we developed. The logic in Figure 2 can be described as follows:

> *An increase in the extent to which NWS contributes to the public's understanding of weather dangers should lead to (1) an increase in the percentage of the public that have emergency preparedness kits, (2) an increase in the percentage of people who alter their behavior after hearing a warning, and (3) the percentage of people who seek shelter as a first action after hearing a warning. The combined result of these influences on the publis should be to reduce the number of injuries related to weather events.*

Finally, Table 13 summarizes the data that were collected for each potential metric by event type to demonstrate the flow depicted in Figure 2.

Table 12 – Summary of Proposed Metrics, All Categories

Metric	Data Source	Question in Survey/Data Element	Calculation
Better Informed			
Public rating of the extent to which NWS contributes to its understanding of weather dangers *(Collected for different weather event types separately)*	NWS website survey, collected once per year over a three-month period	How would you rate the extent to which the information provided by NWS has contributed to your understanding of the dangers of {EVENT TYPE}-related events on a scale of 1 to 10, where 1 means "not at all" and 10 means "significantly"?	Average score from survey where the score is calculated for each respondent as the respondent's rating minus one divided by 9 and then multiplied by 100.
Get Prepared			
Percent of public that has an emergency preparedness kit *(Collected for different weather event types separately)*	NWS website survey, collected once per year over a three-month period	Do you have a {EVENT TYPE} emergency preparedness kit for your vehicle? (Yes/No)".	Percentage that answer yes to the question.
Take Action			
Percent of public that altered their previous activities after hearing a warning or who remained vigilant *(Collected for different weather event types separately)*	Survey – online panels implemented for a random selection of warnings issued	Did you take any of the following actions when you received the {EVENT TYPE} (advisory/warning)? Did you continue previous activities? Did you monitor weather forecasts? [a]	Percentage that answer **no** to the question on continuing previous activities or who answer **yes** to the question on monitoring weather forecasts.

Metric	Data Source	Question in Survey/Data Element	Calculation
Percent of public that sought protective shelter as a first action *(Collected for different weather event types separately)*	Survey – online panels implemented for a random selection of warnings issued	Which of the following actions was the first action you took when you received the {EVENT TYPE} warning? Response option: seek shelter [b]	Percentage that select "seek shelter" (or option determined to be seeking shelter).
Avoided Injuries			
Reduced number of injuries compared to expected number *(Estimated for different weather event types separately)*	Statistical modeling using NWS StormDat data combined with Census data	Number of injuries reported in StormDat by WFOs	Predictions from a baseline statistical model compared to actual number that occur

[a] The second question about monitoring weather forecasts is not currently part of the questionnaire; ERG is recommending including this question in future surveys.

[b] For extreme heat, the wording should be "seek air conditioning". For winter weather, the wording should be "altered your routine."

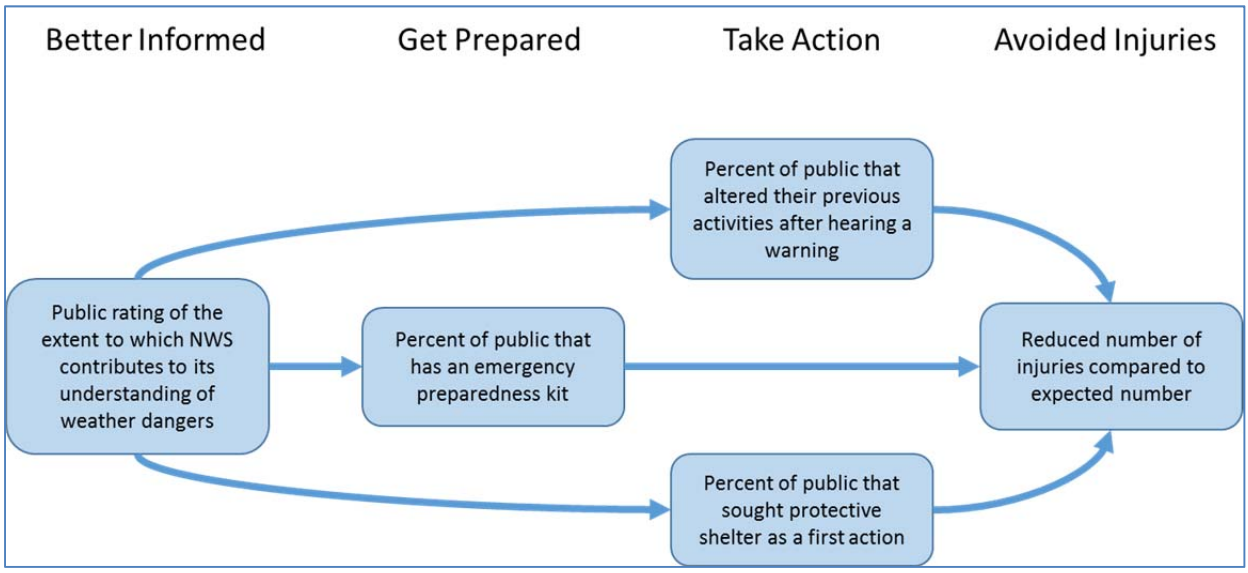

Figure 2 – Logic Model Using Proposed Metrics

Table 13 – Summary of Values for Proposed Metrics by Weather Event Type from Pilot Data

Category	Metric	Winter Weather	Flash Flood	Severe Thunder-storm	Extreme Heat
Better informed	Public rating of the extent to which NWS contributes to its understanding of weather dangers	2015: 80 [a] 2016: 82 [a]	2015: 76 [a] 2016: 75 [a]	[b]	2015: 63 [a]
Get prepared	Percent of public that has an emergency preparedness kit	2015: 57% [a] 2016: 59% [a]	2015: 31% [a] 2016: 23% [a]	[b]	[b]
Take action	Percent of public that altered their previous activities after hearing a warning [c]	51%	29%	35%	41%
Take action	Percentage of public that sought protective shelter as a first action	24%	10%	25%	40%
Avoided Injuries	Number of avoided injuries (2012-2013)	56	7	53	[d]

[a] These values reflect the website survey results, as we recommended for the "better informed" and "take action" metrics.
[b] Not collected for this event.
[c] The survey question asked respondents if they had continued their previous activities when they heard the warning. The percentages here are calculated by subtracting the percentage who did not alter their previous activities from one.
[d] Avoided injuries we not estimated for extreme heat since the statistical model to estimate the baseline model could not be reliably estimated.

6.2 RECOMMENDATIONS

This section provides a set of recommendation for NWS to consider in moving forward with developing societal outcome measures.

1. **Collect data on the "better informed," "get prepared," and "take action" metrics, but not the "avoided injuries" metric.** We expect the data we collected for these three metric categories will provide useful information for NWS in tracking societal outcomes. We are not recommending the one metric category that gets closest to a full societal outcome (avoided injuries). Nevertheless, if the logic model is valid, increases in those who take action should lead to reduced injuries.

2. **Focus on the four weather event types that we have focused on in this report, tracking data for each type separately.** The four events include winter weather, flash floods, severe thunderstorms, and extreme heat. Tracking outcomes from these four event types should provide NWS with an indication of the potential societal outcomes. As can be seen in Table 14, the four events we focused on accounted for 46 percent of injuries and 50 percent of fatalities in 2015 excluding tornadoes from severe weather events; including tornadoes increases those percentages to 89 percent for injuries and 57 percent for fatalities.

Table 14 - Injuries and Fatalities from Weather Events, 2015

Event Type	Injuries		Fatalities	
	Number	Percent	Number	Percent
Severe weather (convective events)				
Lightning	130	6.1%	27	5.2%
Tornadoes	924	43.1%	36	6.9%
Thunderstorm winds	159	7.4%	41	7.9%
Extreme Heat	640	29.9%	45	8.6%
Flash Floods	42	2.0%	129	24.7%
Winter Weather	17	0.8%	20	3.8%
Four focal events	**1912**	**89.2%**	**298**	**57.1%**
Four focal events, excluding tornadoes	**988**	**46.1%**	**262**	**50.2%**
All	**2,143**	**-**	**522**	**-**

Source: http://www.nws.noaa.gov/om/hazstats/sum15.pdf.

3. **For the "better informed" metric, ERG recommends tracking the metric we referred to as "Public rating of the extent to which NWS contributes to its understanding of weather dangers."** To collect data for this metric, NWS should:
    a. Use the question on the extent to which respondents felt NWS contributed to their understanding of the specific weather event.
    b. Collect these data from the website pop-up survey during one quarter of the year.
    c. Translate the values to a metric using the calculation currently used by CFI Group (see Section 2.2).
    d. Review the demographics for the samples of each quarter to assess whether changes in the metric values are attributable to changes in demographics rather than change in the associated outcome.

4. **For the "get prepared" metric, ERG recommends tracking the metric we referred to as "Percent of public that has an emergency preparedness kit."** To collect these data, NWS should:
    a. Use the question on whether or not the respondents had emergency kits.
    b. Collect these data from the website pop-up survey during one quarter of the year.
    c. Review the demographics for the samples of each quarter to assess whether changes in the metric values are attributable to changes in demographics rather than change in the associated outcome.

5. **For the "take action" metric, ERG recommends tracking two metrics: "Percent of public that altered their previous activities after hearing a warning or who remain vigilant" and "Percentage of public that sought protective shelter as a first action."** To collect these data, NWS should follow the random event (NWS warning) selection process outlined in Section 4.4.2. NWS should also assess whether where the data were collected impacted the values of the

metric. For example, collecting these data for winter weather in New England, will result in a set of respondents who are familiar with how to deal with winter storms; comparing those results to results from areas less familiar with winter weather events may skew the results.

6. **NWS should perform additional research into ERG's proposed simplified logic model.** Specifically, the logic model we developed is based on our qualitative research into how societal outcomes are generated. The revised version of the model in Figure 2 proposes links between the metrics we developed. Further quantitative research is needed to validate the connections in both the original simplified model (Figure 1) and the refined model (Figure 2).

7. **NWS should perform further research into the avoided injuries modeling approach.** ERG expended as much project resources as possible on this area and we had some success in developing estimates. However, ERG had to limit the resources in this area to allow for collection and analysis of data for other metrics. Additional research is warranted since ERG was able to develop some initial valid models. Further research could help refine and validate the models already developed. One area of potential improvement would be adding in measures of severity for each event type.