Evaluation of Real-Time Mortality Surveillance Based on Media Reports

Dr Olaniyi O. Olayinka, MD, MPH,
Centers for Disease Control and Prevention, Health Studies Branch, Atlanta, Georgia

Dr Tesfaye M. Bayleyegn, MD,
Centers for Disease Control and Prevention, Health Studies Branch, Atlanta, Georgia

Ms Rebecca S. Noe, MN, MPH, FNP-BC,
Centers for Disease Control and Prevention, Health Studies Branch, Atlanta, Georgia

Dr Lauren S. Lewis, MD,
Centers for Disease Control and Prevention, Health Studies Branch, Atlanta, Georgia

Ms Vincent Arrisi, MD, and
New Jersey Department of Health, Office of Vital Statistics and Registry, Trenton, New Jersey

Amy F. Wolkin, DrPH, MSPH
Centers for Disease Control and Prevention, Health Studies Branch, Atlanta, Georgia

Abstract

Objective—We evaluated the usefulness and accuracy of media-reported data for active disaster-related mortality surveillance.

Methods—From October 29 through November 5, 2012, epidemiologists from the Centers for Disease Control and Prevention (CDC) tracked online media reports for Hurricane Sandy–related deaths by use of a keyword search. To evaluate the media-reported data, vital statistics records of Sandy-related deaths were compared to corresponding media-reported deaths and assessed for percentage match. Sensitivity, positive predictive value (PPV), and timeliness of the media reports for detecting Sandy-related deaths were calculated.

Results—Ninety-nine media-reported deaths were identified and compared with the 90 vital statistics death records sent to the CDC by New York City (NYC) and the 5 states that agreed to participate in this study. Seventy-five (76%) of the media reports matched with vital statistics records. Only NYC was able to actively track Sandy-related deaths during the event. Moderate sensitivity (83%) and PPV (83%) were calculated for the matching media-reported deaths for NYC.

Conclusions—During Hurricane Sandy, the media-reported information was moderately sensitive, and percentage match with vital statistics records was also moderate. The results indicate that online media-reported deaths can be useful as a supplemental source of information for
situational awareness and immediate public health decision-making during the initial response stage of a disaster.

**Keywords**
public health surveillance; disasters; hurricane; media

In the North Atlantic Ocean, hurricanes typically occur from June 1 to November 30. Hurricanes have circulating winds of at least 74 miles per hour and can cause significant loss to human lives and damage to property. The frequency and intensity of hurricanes are predicted to rise over the next 3 decades with attendant consequences on the public’s health.

Human deaths are a reliable marker of the human impact of a disaster and an emergency. Mortality data can reveal geographical clustering of deaths, circumstances, and risk factors for deaths during a disaster. This epidemiologic information allows specific public health interventions to be efficiently targeted at the right populations. Hence, collecting timely and accurate disaster-related mortality data during hurricanes and other disasters is critical for immediate public health action. Obtaining disaster-related mortality data, however, is challenging. To decrease hurricane-related morbidity and mortality, public health emergency response professionals need to leverage existing public health resources for evidence-driven decision-making.

On October 29, 2012, Hurricane Sandy hit the northeastern US coastline at New Jersey. High winds and heavy rainfalls from the category 2 storm caused significant damage in New York City (NYC) and the states of Connecticut, Maryland, New Hampshire, North Carolina, New York, Pennsylvania, Virginia, West Virginia, and New Jersey. The storm surge precipitated by Hurricane Sandy led to severe flooding and power outages, which added to the destructive impact of this event. The economic loss exceeded $68 billion in damages, making Hurricane Sandy the second costliest storm in US history. The public health impact of Hurricane Sandy included injuries, illness, and deaths.

During the response to Hurricane Sandy, the Centers for Disease Control and Prevention (CDC) tested the use of online media reports for disaster-related mortality surveillance. The CDC used the Google search engine (Google Inc, Mountain View, CA) to identify deaths reported on the Internet (eg, websites of traditional news media and blogs).

The data were used to characterize decedent demographics; circumstance, cause, and location of death; and associated risk factors. The disaster epidemiology response team shared this information with the CDC emergency operation center for situational awareness and possible public health action. The objective of this pilot study was to evaluate the information captured by the media reports and determine the usefulness and accuracy of media-reported data for active disaster-related mortality surveillance in the future.
METHODS

From October 29 through November 5, 2012, CDC epidemiologists searched the Internet each day for reports of Sandy-related deaths by using the keywords “death,” “disaster,” “drowning,” “Hurricane Sandy,” “memorial,” “Sandy,” or “storm.” The tracking was conducted for 1 week because the CDC emergency operation center was only active for that duration. For each media-reported death, the CDC attempted to identify the decedent’s name(s), age, sex, cause of death, circumstance of death, place of death, date of death, and the webpage and memorial sites providing the information (eg, CNN, New York Times, Associated Press, Legacy.com). Duplicates were removed from the list.

To evaluate the data obtained from the media reports, CDC contacted vital statistics offices in the affected states 1 year after the event to request Sandy-related vital statistics records. States were asked to identify Sandy-related deaths that occurred from October 29 through November 5, 2012, by using the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) diagnoses codes X37 (cataclysmic storm) or X37.0 (hurricane), or by conducting text string searches using keywords such as “Hurricane Sandy,” “storm related,” or “drowning.” Sandy-related deaths other than those caused by drowning were also captured in our online search.

To determine the usefulness and accuracy of the media reports, the CDC guidelines for evaluating public health systems were adopted while the vital statistics record served as the gold standard for this evaluation. The usefulness of the media reports was defined on the basis of the report’s ability to detect a Sandy-related death and report the demographic information, circumstance under which a death occurred, and the location of death during the initial response stage of Hurricane Sandy. In addition, we considered the accessibility and availability of the media-reported data to support public health decision-making. Accuracy (ie, data quality) was assessed on the basis of the completeness and validity of the media reports. For the purpose of this evaluation, a match of media-reported data with vital statistics records of 70% to 85% was considered to be moderate.

We compared information from the media-reported deaths with the information from the matching vital statistics records from the participating states. We matched media-reported deaths by using 2 matching categories. First, we matched the media records with their corresponding vital statistics record by using all of the following key attributes: decedent’s (1) first and (2) last names (minor spelling errors accepted), (3) place of death (ie, state, city, county, borough, or specific place of death), and (4) circumstance or cause of death. Second, for the media records with names missing, we matched with the vital statistics record the frequency of the following key variables that were reported by the media: name (first and last), sex, age, place of death, circumstances of death, date of death, and cause of death.

In the jurisdiction where Sandy-deaths could be identified independently without reference to our list (ie, New York City), we calculated sensitivity and positive predictive value (PPV) and determined how quickly deaths related to Sandy were first reported (timeliness). Deaths that were reported in the media to have occurred outside the study period of October 29
through November 5, 2012, as well as those outside the 5 states that agreed to participate in this pilot study and NYC were excluded.

RESULTS

Through the Internet searches, the CDC identified 99 media-reported deaths from the 5 states (Connecticut, North Carolina, New Jersey, Pennsylvania, and Virginia) and NYC that agreed to participate in this evaluation (Figure 1); an additional 16 media-reported deaths were identified from states that were not included in this study. Of the 99 media records, 100% had information on circumstance of death, 95% had information on the sex of decedents, 94% had information on the age of decedents, and 79% had information on the name of decedents (Figure 2). Except NYC, the other vital statistics offices were not able to identify Sandy-related deaths by using the ICD-10-CM diagnoses codes X37 or X37.0 or by conducting text string searches by using the keywords CDC provided. NYC was the only jurisdiction able to identify their Sandy-related deaths (41 [41.5%]) by using the ICD-10-CM diagnosis codes or keyword search. The 5 participating states used the list of Sandy-related deaths the CDC had obtained from online media reports to identify the corresponding vital statistics records of their Sandy-related deaths (49 [49.5%]). In total, the CDC received 90 Hurricane Sandy–related vital statistics records from NYC and the 5 states; the CDC attempted to match those records with the 99 media records (Figure 3).

Of the 99 media records, 6 media-reported deaths (6%) from NYC actually occurred outside their jurisdiction, the death records of 2 deaths (2%) were not provided, and 1 death (1%) occurred outside the study period (ie, October 29–November 5, 2012). All 9 were deaths reported by the media for NYC. Of the remaining 90 media records, 75 (76%) matched with their corresponding vital statistics records by use of most of the key attributes, ie, decedents’ first and last names, place of death, and circumstance or cause of death (Table 1). Fifteen media-reported deaths (15%) were not matched and had inadequate information to allow a match to a vital statistics record by use of our criteria. Of the 75 matched records, 60 (61%) matched on name (first or last name), place, and circumstance or cause of death; the remaining 15 (15%) matched on sex, place, and circumstance or cause of death.

NYC was the only jurisdiction that identified Sandy-related deaths without the CDC-provided list of decedents. Therefore, we only calculated sensitivity and PPV of the media reports for NYC and requested from them how quickly they captured and reported Sandy-related deaths during the disaster. The sensitivity and PPV were each 83%, with 34 of the media-reported deaths found to be a match among the 41 deaths recorded through NYC vital statistics. NYC reported Sandy-related deaths within 24 hours by using their vital statistics records and electronic death registration systems (eDRS). Similarly, the CDC observed that online media reports of Sandy-related deaths were updated daily.

DISCUSSION

Collecting timely and accurate disaster-related mortality data is challenging during a hurricane or other disaster but critical to the public health emergency response.7,19 In a 2012 Council of State and Territorial Epidemiologists disaster capacity survey of disaster
epidemiology and surveillance practices of 53 state and territorial public health departments, only 33% of jurisdictions indicated that their existing mortality surveillance system proved adequate during a past disaster. Additionally, existing vital statistics systems have time lags inherent in processing death certificates. The adoption of eDRS by some states is believed to increase the timeliness of data collection; however, not all states have adopted eDRS. Another difficulty is that many death certificates do not indicate disaster-relatedness under cause of death. To quickly bridge such gaps in disaster-related mortality surveillance, alternative, supplemental data sources (eg, media reports of disaster-related health impacts) could be used to improve situational awareness and support immediate public health decision-making, especially during the initial phase following a disaster. For instance, media reports could be used to overcome the time lag in reporting of disaster-related deaths by traditional surveillance systems and provide public health professionals access to readily available information during the initial phase of a disaster.

Timely information on the number of disaster-related deaths is a basic metric of the event’s impact on a population. Online obituary services and online media reports of deaths are potential resources for tracking disaster-related mortality nationwide. The use of novel data sources, such as media reports, could enhance public health surveillance. Concerns about influenza and foodborne illness outbreaks have already prompted some state health departments and researchers to consider media reports as a data source. For example, NYC health officials recently tracked comments about restaurants appearing on Yelp (San Francisco, CA), a business review website, as a potential resource to discover unreported food-borne illness outbreaks. Further, an effort to visualize and aggregate media reports for influenza led to the creation of HealthMap, an online, interactive system that uses freely accessible, real-time health information to map emerging public health threats. Similarly, a free web-based surveillance tool for influenza, Google Flu Trends (Google Inc, Mountain View, CA), has been shown to detect influenza outbreaks approximately 1 week before national influenza data are reported.

Another important attribute of online media reports is timeliness and the ability to identify Hurricane Sandy–related deaths. During the pilot study, we observed that online media reports were frequently updated. In this evaluation, NYC reported actively tracking Sandy-related deaths during the response for daily situational awareness reporting. The average time it took the media to report deaths was 24 hours; NYC was also able to report Sandy-related deaths within 24 hours by using their vital statistics records and eDRS. NYC used its eDRS to track deaths in real time. NYC has a centralized medical examiner system and protocols for attributing hurricane as cause of death; hence, their public health officials were able to access preliminary death certificate data, flag disaster-related deaths, and search cause of death text fields for hurricane-related terms during the response.

However, because none of the states initially provided death records of Sandy-related deaths using those methods, we provided them with a list we had obtained from online media reports that contained the number of deaths, the decedent’s name(s), county, city and location of death, date of death or body recovery, age, sex, and possible cause and circumstance of death to facilitate the search for the records by using single or combined variables of the list provided.
Unlike NYC, many states, including Connecticut, North Carolina, Pennsylvania, and Virginia, do not have eDRS or do not have the ability to leverage eDRS for "near to real-time" disaster mortality surveillance.\textsuperscript{29,30} For states that cannot use eDRS for disaster mortality surveillance, death certification and registration can be time-intensive and delay public health emergency response.\textsuperscript{9,31} Some states do initiate active mortality surveillance systems during an event to improve timeliness. In 2007, for example, Texas created a standalone active disaster-related mortality surveillance system (DRMS). The system is activated immediately after a disaster and maintained by the collaborative effort of local and regional health departments, medical examiners, and justices of the peace. Texas DRMS is used to track disaster-related deaths for a minimum of 6 weeks after a disaster.\textsuperscript{10}

The findings from this evaluation showed that online media reports provided moderately accurate and useful mortality data during Hurricane Sandy. Most of the media records matched with their corresponding vital statistics records by use of most of the key attributes. In addition, the 2 CDC epidemiologists who tracked Sandy-related deaths were able to conduct an online search of media reports twice daily (ie, one when resuming to work in the morning and the other before close of work, each spending an average of 2 hours daily for the search). The media reports were disseminated to the CDC emergency operations center, who used this information for situational awareness during Hurricane Sandy.

Evaluation of online media reports for NYC was similar to the results of a study showing that \textit{Pittsburgh Post-Gazette} Internet death notices provided accurate, timely mortality data for Allegheny County, Pennsylvania, deaths compared with all mortality records from the Pennsylvania Department of Health for 1998 to 2001.\textsuperscript{21} In our evaluation, the overall match was moderate (76%), although it was difficult to conduct an initial comparison of the media reports with vital statistics record by cause of death. This was because of missing information rather than incorrect information. Cause of death is defined as the disease or injury that initiated the train of morbid events (the underlying cause) leading directly (the immediate cause) to death or the circumstance of the accident or violence that produced the fatal injury.\textsuperscript{32} The media often do not report cause of death (eg, blunt force trauma); however, cause of death is an important feature of a vital statistics record. Although media reports might lack cause of death, they typically include circumstance of death. This field is also on the vital statistics record and captures the manner of the death (eg, accident, suicide, homicide) and how the death occurred (eg, car accident).\textsuperscript{32} Assessments of these media-reported circumstances can be used to prioritize urgent public health messaging before vital statistics records are available.

The results of this evaluation are subject to several limitations. The key words used in conducting online text searches of Hurricane Sandy–related deaths were not exhaustive. For example, we did not include the key word “superstorm” in our online search and might have missed some media reports of deaths attributed to Superstorm Sandy. Tracking disaster-related deaths using keywords is relatively slow compared with computer algorithms or interfaces that automatically scan the Internet. Future works may consider developing such automated algorithms that could be used for active surveillance of disaster-related deaths. Our analysis might have missed some Hurricane Sandy–related deaths that were reported through vital statistics. In addition, we were only able to calculate sensitivity and PPV for
the media reports for NYC and not for the other 5 participating states. Therefore, we were unable to determine the true number of deaths attributable to Hurricane Sandy from the 5 states. Finally, 15 media-reported Hurricane Sandy–related deaths were not included in the analysis because they had inadequate information to allow a match to a vital statistics record using our criteria. These might bias the outcome of our analysis in favor of the media reports. A strength of this study was the use of vital statistics records from all jurisdictions affected by Hurricane Sandy to validate the online media reports of deaths attributed to the storm.

CONCLUSIONS

During Hurricane Sandy, the online media reported more Sandy-related deaths than vital statistics and provided timely information on Hurricane Sandy–related deaths. Compared with vital statistics records in NYC, the media-reported information was moderately sensitive and many of the key attributes matched the vital statistics record. Notably, 5 states were unable to identify Hurricane Sandy–related deaths through vital records by use of key words or text string searches. Although formal reporting and surveillance remain essential to protect public health, online media-reported deaths can supplement available information on disaster-related deaths for situational awareness and immediate public health decision-making during the initial response stage of a disaster. Use of traditional sources of information, such as vital statistics or medical examiner records, might be warranted when more accurate information is needed. However, if deaths are not actively tracked during a disaster, it may be difficult or impossible to retrospectively identify those deaths as being disaster-related in jurisdictions where active disaster-related mortality surveillance is not conducted.

Acknowledgments

We thank Ann Madsen of the Office of Vital Statistics, New York City Department of Health and Mental Hygiene; Kathleen Jones-Vessey of North Carolina State Center for Health Statistics, North Carolina Department of Health and Human Services; Jane Purtill of the Office of Vital Records, Connecticut Department of Public Health; Janet Rainey of the Division of Vital Records, Commonwealth Virginia Department of Health; and Diane Kirsch and James Rubertone of the Division of Statistical Registries, Pennsylvania Department of Health, for their support for this project.

References


FIGURE 1.
Path of Hurricane Sandy, Impact, and Hurricane-Related Deaths Reported by the Media, by County: October 29–November 5, 2012.
FIGURE 2.
99 Media-reported deaths identified

Death records of the 99 media-reported deaths were requested from the offices of vital statistics (CT, NC, NJ, NYC, PA, VA)

90 deaths records were received for comparison

9 media-reported deaths occurred outside the study area (n=6) and study period (n=1), and their death records were not provided (n=2).

15 media-reported had insufficient information for matching

75 Vital statistics records matched with Media-reported deaths
• Match by Name(s), Place & Circumstance or Cause Death (n=60)
• Match by Sex, Place & Circumstance or Cause of Death (n=15)

FIGURE 3.
### TABLE 1


<table>
<thead>
<tr>
<th></th>
<th>CT, No. (%)</th>
<th>NC, No. (%)</th>
<th>NJ, No. (%)</th>
<th>NYC, No. (%)</th>
<th>PA, No. (%)</th>
<th>VA, No. (%)</th>
<th>Total No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media reports</td>
<td>5 (5)</td>
<td>4 (4)</td>
<td>24 (24)</td>
<td>47 (47)</td>
<td>17 (17)</td>
<td>2 (2)</td>
<td>99</td>
</tr>
<tr>
<td>Vital statistics</td>
<td>5 (6)</td>
<td>2 (2)</td>
<td>23 (26)</td>
<td>41 (46)</td>
<td>17 (19)</td>
<td>2 (2)</td>
<td>90</td>
</tr>
<tr>
<td>Matched by: name, place AND circumstance or cause of death</td>
<td>5 (100)</td>
<td>2 (50)</td>
<td>14 (58)</td>
<td>26 (55)</td>
<td>11 (65)</td>
<td>2 (100)</td>
<td>60 (61)</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match overall</td>
<td>5 (100)</td>
<td>2 (50)</td>
<td>19 (79)</td>
<td>34 (72)</td>
<td>13 (76)</td>
<td>2 (100)</td>
<td>75 (76)</td>
</tr>
<tr>
<td>Unmatched</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>5 (21)</td>
<td>8 (17)</td>
<td>2 (12)</td>
<td>0 (0)</td>
<td>15 (15)</td>
</tr>
</tbody>
</table>