A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP6976.

Assessing United States County-Level Exposure for Research on Tropical Cyclones and Human Health

G. Brooke Anderson,[1] Joshua Ferreri,[1,2] Mohammad Al-Hamdan,[3] William Crosson,[3] Andrea Schumacher,[4] Seth Guikema,[5] Steven Quiring,[6] Dirk Eddelbuettel,[7] Meilin Yan,[1,8] and Roger D. Peng[9]

[1]Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, Colorado, USA
[2]School of Medicine, University of Colorado, Aurora, Colorado, USA
[3]Universities Space Research Association, NASA Marshall Space Flight Center, Huntsville, Alabama, USA
[4]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado, USA
[5]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA
[6]Department of Geography, Ohio State University, Columbus, Ohio, USA
[7]Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA
[8]Beijing Innovation Center for Engineering Science and Advanced Technology and State Key Joint Laboratory of Environmental Simulation and Pollution Control, Peking University, Beijing, China
[9]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Background: Tropical cyclone epidemiology can be advanced through exposure assessment methods that are comprehensive and consistent across space and time, as these facilitate multiyear, multistorm studies. Further, an understanding of patterns in and between exposure metrics that are based on specific hazards of the storm can help in designing tropical cyclone epidemiological research.

Objectives: *a*) Provide an open-source data set for tropical cyclone exposure assessment for epidemiological research; and *b*) investigate patterns and agreement between county-level assessments of tropical cyclone exposure based on different storm hazards.

Methods: We created an open-source data set with data at the county level on exposure to four tropical cyclone hazards: peak sustained wind, rainfall, flooding, and tornadoes. The data cover all eastern U.S. counties for all land-falling or near-land Atlantic basin storms, covering 1996–2011 for all metrics and up to 1988–2018 for specific metrics. We validated measurements against other data sources and investigated patterns and agreement among binary exposure classifications based on these metrics, as well as compared them to use of distance from the storm's track, which has been used as a proxy for exposure in some epidemiological studies.

Results: Our open-source data set was typically consistent with data from other sources, and we present and discuss areas of disagreement and other caveats. Over the study period and area, tropical cyclones typically brought different hazards to different counties. Therefore, when comparing exposure assessment between different hazard-specific metrics, agreement was usually low, as it also was when comparing exposure assessment based on a distance-based proxy measurement and any of the hazard-specific metrics.

Discussion: Our results provide a multihazard data set that can be leveraged for epidemiological research on tropical cyclones, as well as insights that can inform the design and analysis for tropical cyclone epidemiological research. https://doi.org/10.1289/EHP6976

Introduction

Hurricanes and other tropical cyclones can severely impact human health in U.S. communities, as shown in several long-term national studies (Rappaport 2000, 2014; Rappaport and Blanchard 2016; Czajkowski and Kennedy 2010; Czajkowski et al. 2011; Moore and Dixon 2012). These studies characterize tropical cyclone health impacts using fatality data aggregated in large part from disaster-related mortality ascertainment and surveillance conducted by agencies like the U.S. Centers for Disease Control and Prevention, the U.S. National Weather Service, and local vital statistics departments. These fatality data are based on case-by-case ascertainment: identifying and characterizing deaths for which there is a clear link with a disaster through death certificate coding or other indicators (Rocha et al. 2017).

For several types of climate-related disasters—including heat waves, floods, and wildfires—such research has been richly supplemented by studies that estimate community-wide excess mortality and morbidity associated with the disaster (e.g., Anderson and Bell 2011; Son et al. 2012; Haikerwal et al. 2015; Liu et al. 2017; Milojevic et al. 2017). These studies, which typically use time-series or case-crossover designs, compare the community-wide rate of a health outcome during a disaster to rates during comparable nondisaster periods. Although such studies cannot attribute specific cases (e.g., a specific death) to a disaster, they can quantify the community-wide change in health risk during disasters and capture impacts that might be missed or underestimated with traditional disaster-related mortality ascertainment and surveillance. In many cases, these studies analyze multiyear, multicommunity data, allowing them to estimate average associations over many disasters and to explore how a disaster's characteristics, or the characteristics of the affected communities, modify associated health risks (e.g., Anderson and Bell 2011; Son et al. 2012; Liu et al. 2017). Some studies have begun to use this approach to study the health impacts of tropical cyclones, including several studies of Hurricane Maria (e.g., Santos-Lozada and Howard 2018; Santos-Burgoa et al. 2018), Hurricane Sandy (e.g., Kim et al. 2016; Mongin et al. 2017; Swerdel et al. 2014), and the 2004 hurricane season in Florida (McKinney et al. 2011). However, to expand this approach to longer time periods and larger sets of communities, researchers must be able to assess exposure to tropical cyclones consistently and comparably across storms, years, and communities.

The National Hurricane Center (NHC) publishes a "Best Tracks" data set that is considered the gold standard for Atlantic-basin tropical cyclone tracks. It records the central position of a tropical cyclone every 6 h, as well as the storm's minimum

Address correspondence to G. Brooke Anderson, Department of Environmental & Radiological Health Sciences, Colorado State University, 1681 Campus Delivery, Fort Collins, CO 80523 USA. Email: brooke.anderson@colostate.edu

Supplemental Material is available online (https://doi.org/10.1289/EHP6976).

The authors declare they have no actual or potential competing financial interests.

Received 24 February 2020; Revised 22 September 2020; Accepted 25 September 2020; Published 28 October 2020.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

central pressure and maximum sustained surface wind. These data are openly available through the revised Atlantic hurricane database (HURDAT2), a poststorm assessment that incorporates data from several sources, including satellite data and, when available, aircraft reconnaissance data (Landsea and Franklin 2013; Jarvinen and Caso 1978). With these data, it is straightforward to measure a storm's direct path and so to measure whether a community was on or within a certain distance of that central track. Some previous epidemiological studies have done this to assess exposure to a tropical storm (e.g., Currie and Rossin-Slater 2013; Kinney et al. 2008; Caillouët et al. 2008), and this method can be applied consistently across years and communities for large-scale studies.

However, when epidemiological studies assess exposure to a tropical cyclone using, as a proxy, how close the storm came to a community, they may misclassify exposure (Grabich et al. 2015a). Although a number of tropical cyclone hazards are strongly associated with distance from the tropical cyclone's center (e.g., wind and, at the coast, storm surge and waves; Rappaport 2000; Kruk et al. 2010), other hazards like heavy rainfall, floods, and tornadoes can occur well away from the tropical cyclone's central track (Rappaport 2000; Atallah et al. 2007; Moore and Dixon 2012). For example, tornadoes generated by tropical cyclones, which were linked to more than 300 deaths in the United States between 1995 and 2009, most often occur 200–500 km from the tropical cyclone's center (Moore and Dixon 2012).

Further, when studies use distance from the storm's track to assess exposure, they often use the same buffer constraints on each side of the storm track (e.g., Kinney et al. 2008; Currie and Rossin-Slater 2013). However, the forces of a tropical cyclone tend to be distributed around its center nonsymmetrically. In the Northern Hemisphere, cyclonic winds are counterclockwise, so extreme winds are more common to the track's right, where counterclockwise cyclonic winds move in concert with the tropical cyclone's forward motion (Halverson 2015), and almost all of the fatal tornadoes associated with U.S. tropical cyclones between 1995 and 2009 occurred to the right of the tropical cyclone's track (Moore and Dixon 2012). Rain, conversely, is often heaviest to the left of the track, especially when the tropical cyclone interacts with other weather systems (Atallah and Bosart 2003; Atallah et al. 2007; Zhu and Quiring 2013) or undergoes an extratropical transition (Elsberry 2002).

The multihazard nature of tropical cyclones therefore makes it difficult to assess exposure based on how close the storm's central track came to the community. Although other approaches have been developed to incorporate storm hazards, particularly wind, into exposure assessment (e.g., Grabich et al. 2015a; Zandbergen 2009; Czajkowski et al. 2011), there is not yet a standard approach. When different studies use different data sets or storm hazards when assessing storm exposure, it becomes difficult to compare and aggregate findings.

To help epidemiological researchers assess exposure to tropical cyclones, we developed an open-source data set (Anderson et al. 2020a), which we present here. These data cover all counties in states in the eastern half of the United States over multiple years and include five metrics characterizing exposure to tropical cyclones (Table 1): *a*) closest distance the storm's central track came to the county's center (a proxy measurement of storm exposure used in some previous studies); *b*) peak sustained surface wind at the county's center over the course of the storm; *c*) cumulative rainfall in the county over the course of the storm; *d*) flooding in the county concurrent with the storm; and *e*) tornadoes in the county concurrent with the storm. We aggregated these data at the county level because data for epidemiological studies are often available at this level [e.g., direct hurricane-related deaths (Czajkowski et al. 2011); birth outcomes (Grabich et al. 2015a, 2015b); autism prevalence (Kinney et al. 2008)] and because decisions and policies to prepare for and respond to tropical cyclones are often undertaken at the county level (Zandbergen 2009; Rappaport 2000). We also wrote functions to explore and map these data and to link them with county-level health data (Anderson et al. 2020b).

Table 1. Exposure metrics considered to assess county-level exposure to tropical cyclones.

Metric	Available	Criteria for exposure
Distance	1988–2018	County population mean center within 100 km of storm track
Rain	1988–2011	County cumulative rainfall of ≥75 mm from 2 d before to 1 d after the storm's closest approach and the storm track came within 500 km of the county's population mean center
Wind	1988–2018	Modeled peak sustained surface wind of ≥34 knots at the county's population mean center
Flood	1996–2018	Flood event listed in the NOAA Storm Events database for the county with a start date within 2 d of the storm's closest approach and the storm track came within 500 km of the county's population mean center
Tornado	1988–2018	Tornado event listed in the NOAA Storm Events database for the county with a start date within 2 d of the storm's closest approach and the storm track came within 500 km of the county's population mean center

Here, we describe how we developed this data set and explore how its measurements compare to other data that could be used to characterize tropical cyclone hazards for epidemiological research. Further, we expand on previous research on measuring exposure to tropical cyclones for epidemiological research (Grabich et al. 2015a). This previous study investigated differences in which communities were assessed as exposed to tropical cyclones during the 2004 hurricane season in Florida, comparing exposure assessment based on distance to the storm's track or vs. a metric that incorporated storm-generated winds within the county. They found important differences, concluding that a study may be prone to bias from exposure misclassification if distance to the storm track is used as a proxy for exposure (Grabich et al. 2015a). Here we expand to investigate this question across a larger set of counties and storm seasons. Further, we investigate patterns in exposure classification based on other metrics of storm-related hazards—rainfall, flooding, and tornadoes—important for inland health impacts of tropical cyclones (Czajkowski et al. 2011; Moore and Dixon 2012). These results can help epidemiologists design studies and plan statistical analysis for multiyear, multicommunity studies that estimate excess mortality and morbidity associated with exposure to the hazards brought by tropical cyclones.

Methods

For our study domain, we used all counties in the eastern half of the United States (Figure 1). We included all tropical cyclones between 1988 and 2018 that were tracked in HURDAT2 (Landsea and Franklin 2013) and came within 250 km of at least one eastern U.S. county (Figure 1). We therefore included all land-falling or near-land Atlantic basin tropical cyclones and excluded storms that never neared the United States.

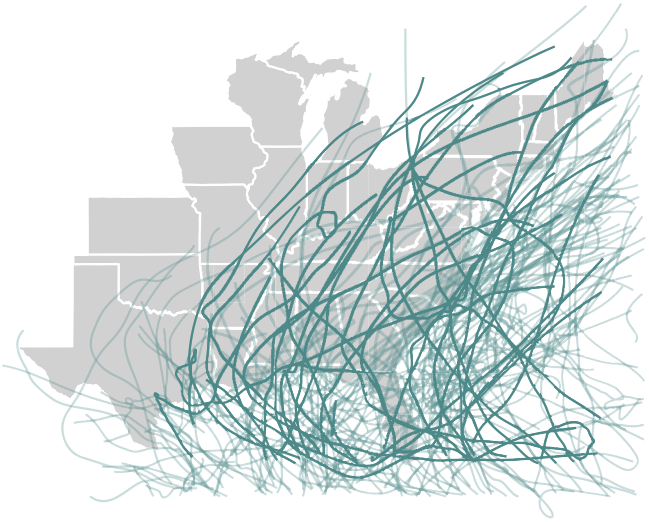

Figure 1. Study area and storms considered in this study. All counties in the states shown in this map were investigated. The lines show the paths of the study storms, which included all tracked storms in 1988–2018 that are recorded in HURDAT2 and that came within 250 km of at least one U.S. county. Thicker lines show the tracks of storms whose names have been retired, indicating that the storm was particularly severe or had notable impacts.

Distance-Based Exposure Metric

We first measured how close each storm's central track came to each county. We used tracking data from HURDAT2, which records the storm center's position at four standardized times for weather data collection (synoptic times), 6:00 A.M., 12:00 P.M., 6:00 P.M., and 12:00 A.M. Coordinated Universal Time (UTC). We interpolated this position to 15-min intervals using natural cubic spline interpolation (Anderson et al. 2020a). At each 15-min interval, we measured the distance between the storm's center and each county's population mean center, as of the 2010 US Decennial Census (U.S. Census Bureau 2020). We took the minimum distance for each county as a measure of how close the storm came to the county over its lifetime. We also recorded the time when this occurred for each county so we could link observed data on precipitation, flooding, and tornadoes. To allow matching with data recorded in local time (e.g., health data), we converted these times from UTC to local time (Anderson and Guo 2016).

Rain-Based Exposure Metric

We based storm rainfall measurements on precipitation data files from the North American Land Data Assimilation System Phase 2 (NLDAS-2) reanalysis data set, which is available for the continental United States (Rui and Mocko 2014). We used data that were previously aggregated from this reanalysis data set to the county level by two of the coauthors (Al-Hamdan and Crosson) for the U.S. Centers for Disease Control and Prevention's Wide-Ranging Online Data for Epidemiological Research (WONDER) database (U.S. CDC 2019; Al-Hamdan et al. 2014). To create these aggregated data, these coauthors took hourly precipitation measurements in the NLDAS-2 precipitation files, which are given at $1/8^{\circ}$ grid points across the continental United States, and summed them by day for each grid point. This approach created a daily precipitation total for each grid point; these were then averaged for all grid points within a county's 1990 U.S. Census boundaries (Al-Hamdan et al. 2014; U.S. CDC 2019). This process generated daily county-level precipitation estimates for each continental U.S. county for 1988–2011 (U.S. CDC 2019). In this study, we matched these county-level daily measurements by date with storm tracks, using the date when the storm was closest to each county. Given the location of storm-affected counties and the typical timing of tropical cyclones, these precipitation measures primarily represent rainfall, although occasionally they may represent snowfall or other types of precipitation.

In the open-source data (Anderson et al. 2020a), we provide precipitation values at a daily resolution for each county for the period from 5 d before to 3 d after each storm's closest approach to the county. In the software that we published in association with this data (Anderson et al. 2020b), we provide functionality to aggregate these daily values to create cumulative precipitation estimates for custom windows around the date of the storm's closest approach to the county. For example, a user could determine precipitation only on the day the storm was closest to each county but could also determine the cumulative rainfall for a more extended period (e.g., 2 d before to 2 d after the storm's closest approach). For the analysis of binary exposure classifications in this study, we calculated storm-associated rainfall as the sum of precipitation from 2 d before the storm's closest approach to the county to 1 d after.

To validate these precipitation estimates, we compared them with ground-based observations in a subset of study counties. We selected nine sample counties geographically spread across storm-prone regions of the eastern United States and for which precipitation data were available from multiple ground-based stations in the Global Historical Climatology Network throughout the study period (Menne et al. 2012; Chamberlain 2017; Severson and Anderson 2016). These sample U.S. counties were: Miami-Dade, Florida; Harris County, Texas; Mobile County, Alabama; Orleans Parish, Louisiana; Fulton County, Georgia; Charleston County, South Carolina; Wake County, North Carolina; Baltimore County, Maryland; and Philadelphia County, Pennsylvania. We summed daily station-specific measurements from 2 d before to 1 d after each storm's closest approach and then averaged these cumulative station-based precipitation totals for each county to create a countywide estimate of cumulative storm-related precipitation based on ground-based monitoring. We measured the rank correlation (Spearman's ρ; Spearman 1904) between storm-specific cumulative precipitation estimates for the two data sources (NLDAS-2 reanalysis data vs. ground-based monitoring) within each sample county.

Wind-Based Exposure Metric

We created a data set of county-level peak sustained surface wind during each storm. To do this, we first modeled each storm's wind field to each county's population mean center (U.S. Census Bureau 2020) at 15-min intervals throughout the course of the storm. We used a double exponential wind model based on results from Willoughby (Willoughby et al. 2006) to model 1-min surface wind (10 m above ground) at each county center, based on inputs of the storm's forward speed, direction, and maximum wind speed (Anderson et al. 2020c). Our model incorporated the asymmetry in wind speeds around the tropical cyclone center that results from the storm's forward movement (Phadke et al. 2003). From these 15-min interval estimates, we identified the peak sustained surface wind in each county over the course of each storm.

To validate these modeled county-level peak wind estimates, we compared them with the postseason wind radii, which have been routinely published since 2004 (Knaff et al. 2016). These radii estimate the maximum distance from the storm's center that winds of a certain intensity extend. They give separate estimates

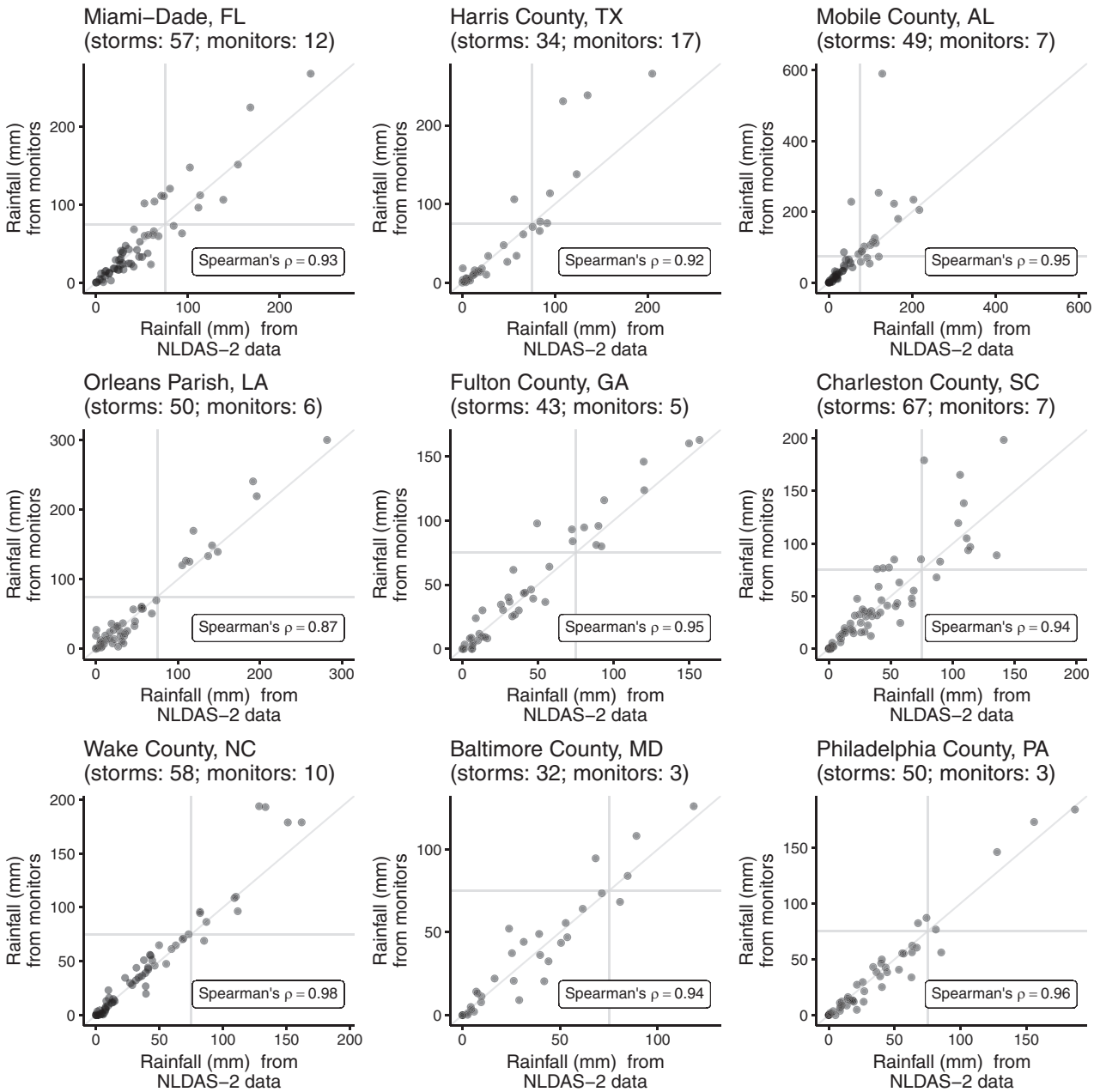

Figure 2. Comparison in nine sample counties of two sources of storm-associated rainfall estimates: (A) county-level estimates derived from a reanalysis data set and (B) county-level estimates based on ground-based observations in nine sample counties. For both, estimates include rainfall from 2 d before to 1 d after the storm's closest approach to the county. Each small graph shows data for one sample county, and each point shows one tropical storm. The number of storms within each county and the number of ground-based monitors reporting precipitation during the county's storms are given above each plot. Horizontal and vertical lines in each plot show the threshold of 75 mm used to classify a storm as "exposed" for the binary classification considered for this metric in further analysis (Table 1). Note that ranges of the x and y axes differ across counties.

for four quadrants of the storm to capture asymmetry in the storm's wind field and are based on a postseason reanalysis that incorporates all available data (e.g., satellite data, aircraft reconnaissance, and ground-based data if available) (Knaff et al. 2016). Wind radii are estimated for three thresholds of peak sustained surface wind: 64, 50, and 34 knots. They therefore allow for the classification of counties into four categories of peak sustained surface wind: <34 knots; 34–49.9 knots; 50–63.9 knots; and ≥64 knots.

We interpolated the wind radii to 15-min increments using linear interpolation (Anderson et al. 2020a) and classified a county as exposed to winds in a given wind speed category if its population mean center was within 85% of the maximum radius for that wind speed in that quadrant of the storm. The 85% adjustment is based on previous research that found that this ratio helps capture average wind extents within a quadrant based on these maximum wind radii (Knaff et al. 2016). We compared these wind radii–based estimates of county-level peak sustained surface wind during a storm with the modeled wind estimates, comparing all study storms since 2004 for which at least one study county experienced peak sustained winds of ≥34 knots (based on the postseason wind radii). For each of these storms, we calculated the percent of study counties that were classified in the same wind category based on both data sources.

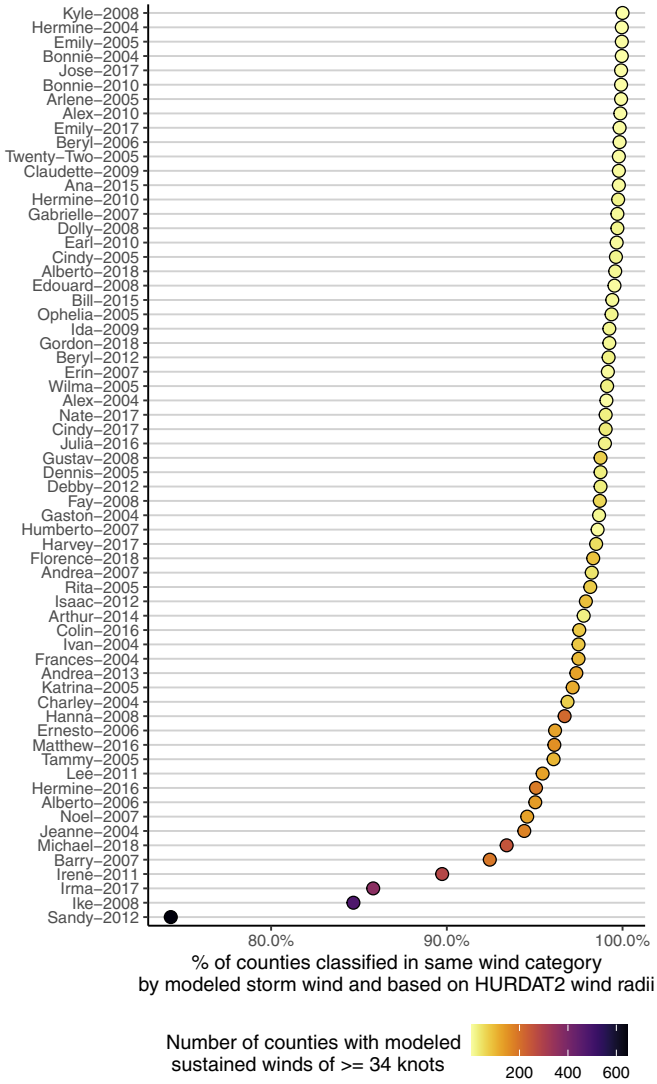

Figure 3. Comparison of two sources of wind exposure estimates: (A) modeled peak sustained surface wind and (B) estimates based on HURDAT2's wind speed radii. Each point represents a storm, with the *x*-axis giving the percent of counties classified in the same category of peak sustained surface wind (<34 knots; 34–49.9 knots; 50–63.9 knots; ≥64 knots) by both sources of data. The color of each point gives the number of study counties that were exposed to peak sustained surface wind of at least 34 knots (based on modeled wind). Estimates are shown for study storms since 2004, the earliest year for which poststorm reanalysis wind speed radii are routinely available in HURDAT2, and for which at least one study county had a peak sustained wind of ≥34 knots based on the poststorm wind radii.

Flood- and Tornado-Based Exposure Metrics

To identify flood- and tornado-based tropical cyclone exposures in U.S. counties, we matched storm tracks with event listings from the Storm Event Database of the National Oceanic and Atmospheric Administration (NOAA/National Weather Service 2020; NOAA NCEI 2020). Events are included in this database based on reports from U.S. National Weather Service personnel and other sources. Although this database has recorded storm data, particularly tornadoes, since 1950, its coverage changed substantially in 1996 to cover more types of storm events, including flood events (NOAA NCEI 2020). We therefore considered only flood metrics of tropical cyclone exposure for storms in 1996 and later.

For each tropical cyclone, we identified all events with event types related to flooding ("Flood," "Flash Flood," "Coastal Flood") or tornadoes ("Tornado") with a start date within a 5-d window centered on the date of the tropical cyclone's closest approach to the county (Anderson et al. 2020a). To exclude events that started near in time to the storm but far from the storm's track and so were likely unrelated to the storm, we excluded any events that occurred ≥500 km from the tropical cyclone's track. "Flood," "Flash Flood," and "Tornado" events in this database were reported by county Federal Information Processing Standard (FIPS) code and so could be directly linked to counties. "Coastal Flood" events were reported by forecast zone; for these, the event was matched to the appropriate county if possible using regular expression matching of listed county names (Anderson and Chen 2019).

The tornado observations from this data set—along with the derived version of the data available through the Storm Prediction Center's National Tornado Database (NOAA/National Weather Service 2020)—are the official tornado event for the United States. There are no other collections of tornado data comparable in temporal or geographic scope, so we did not further validate the tornado event data. It is difficult to characterize flooding at the county level because flooding can be very localized and can be triggered by a variety of causes. To investigate the extent to which the NOAA flood event listings capture extremes that might be identified with other flooding data sources, we investigated a sample of study counties, comparing the flood event data during tropical storms with streamflow measurements at U.S. Geological Survey (USGS) county streamflow gages (USGS 2020; Lammars and Anderson 2017; Hirsch and De Cicco 2015). We considered nine study counties, selecting counties geographically spread through storm-prone areas of the eastern United States and with at least one streamflow gage reporting data during all events. The sample counties were: Baltimore County, Maryland; Bergen County, New Jersey; Escambia County, Florida; Fairfield County, Connecticut; Fulton County, Georgia; Harris County, Texas; Mobile County, Alabama; Montgomery County, Pennsylvania; and Wake County, North Carolina.

For each county, we first identified all streamflow gages in the county with complete data for 1 Jan 1996–31 Dec 2018. If a storm did not come within 500 km of a county, it was excluded from this analysis, but all other study storms were considered. For each storm and county, we summed streamflow measurements across all county gages to generate daily totals for the 5-d window around the storm's closest approach. We took the maximum of these daily streamflow totals as a measure of the county's maximum daily streamflow during that storm. We also calculated the percent of streamflow gages in the county with a daily streamflow that exceeded a threshold of flooding (the streamgage's median value for annual peak flow) on any day during the 5-d window. We investigated how these measurements varied between storms with associated flood events listings vs. storms without an event listing for the county, to explore if storms with flood event listings tended to be associated with higher streamflow at gages within the county.

Binary Storm Exposure Classification

In our open-source data, we provide continuous measurements of some of the exposure metrics: closest distance of each storm to each county, peak sustained surface wind at the county center, and cumulative rainfall. However, epidemiological studies of tropical cyclones often use a binary exposure classification ("exposed" vs. "unexposed") to assess storm-related health risks (e.g., Grabich et al. 2015b; McKinney et al. 2011; Caillouët et al. 2008), so we explored patterns in storm exposure based on binary classifications of these exposure metrics.

Two of the exposure metrics (flood- and tornado-based) were inherently binary in our data, because these metrics were based on whether an event was listed in the NOAA Storm Events database. For other exposure metrics, each county was classified as exposed to a tropical cyclone based on whether the exposure

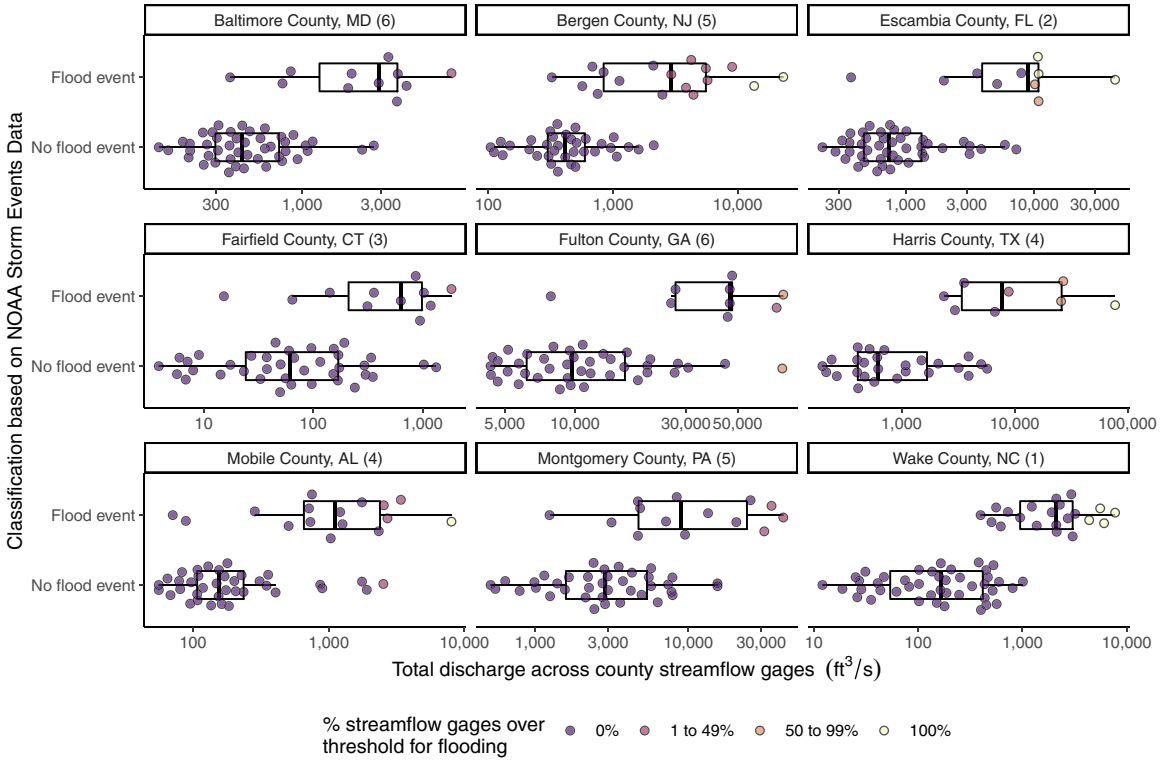

Figure 4. Comparison in nine sample study counties of flood status based on two data sources: (A) NOAA Storm Events listings matched to HURDAT2 storm track data and (B) county streamflow gages. Each small plot shows results for one of the sample counties. Each point represents a single tropical storm, and the point's position along the *x*-axis shows the highest daily total streamflow (cubic feet per second) during the 5-d window surrounding the storm, summed across all identified streamgages in the county. The *y*-axis separates storms for which a flood event was reported in NOAA's Storm Events database in the county with a start date within the 5-d window of the storm's closest approach. The color of each point gives the percent of streamflow gages in the county with a daily streamflow that exceeded a threshold of flooding (the streamgage's median value for annual peak flow) on any day during the 5-d window. The number of streamflow gages used in analysis are given in parentheses beside the county's name in the panel title. Note that the *x*-axis scales differ by county, depending on the number of streamflow gages and typical flow rates for each gage, and are on a log-10 scale.

metric exceeded a certain threshold (Table 1). We picked reasonable thresholds (e.g., the threshold for gale-force wind for wind exposure; Table S1), but others could be used with the open-source data and its associated software.

For the rainfall metric, a distance constraint was also necessary, to ensure that rainfall unrelated and far from the storm track was not misattributed to a storm. Through exploratory analysis, we set this distance metric at 500 km (i.e., for a county to be classified as exposed based on rainfall, the cumulative rainfall in the county had to be over the 75 mm threshold and the storm must have passed within 500 km of the county; Table 1). This distance constraint was typically large enough to capture storm-related rain. However, data users should note that in rare cases—for example, exceptionally large storms (e.g., Hurricane Ike in 2008) or storms for which storm tracking was stopped at extratropical transition (e.g., Tropical Storm Lee in 2011)—some storm-related rain exposures may be missed because of this distance constraint (Figure S1). This distance constraint can be customized using the software published in association with the open-source data (Anderson et al. 2020b).

We characterized patterns in county-level exposure in the eastern United States for exposure assessment based on each of the measured tropical cyclone hazards (cumulative precipitation, peak sustained wind, flooding, and tornadoes). Depending on available exposure data, this assessment included some or all of the period from 1988 to 2018 (Table 1). For each binary metric of exposure to a tropical cyclone hazard, we first summed the total number of county-level exposures over available years and mapped patterns in these exposures.

Finally, we investigated how well exposure assessment agreed across these metrics. These results can help epidemiologists answer several key questions as they interpret previous research and design new studies. Those questions include, for example: If a single storm hazard has been used to measure exposure in an epidemiological study, can the result be interpreted as an association with tropical cyclone exposure in general, or should it be limited to represent an association with that specific storm hazard? If a study investigates the association between a single storm hazard and a health outcome, could this estimate be confounded by other storm hazards? When planning a new study that will incorporate several storm hazards, might there be variance inflation from multicollinearity or difficulties in disentangling the roles of separate hazards?

For each storm we investigated the degree to which the set of counties assessed as exposed based on one metric overlapped with the set assessed as exposed based on each other metric, including a metric that used distance of a county from the storm track as a proxy for exposure to storm hazards. We calculated the within-storm Jaccard index (J_s) (Jaccard 1901, 1908) between each pair of exposure metrics. This approach measures similarity between two metrics ($X_{1,s}$ and $X_{2,s}$) for tropical cyclone s as the proportion of counties in which both of the metrics classify the county as exposed out of all counties classified as exposed by at least one of the metrics:

$$J_s = \frac{X_{1,s} \cap X_{2,s}}{X_{1,s} \cup X_{2,s}}. \quad (1)$$

This metric can range from 0, in the case of no overlap between the counties classified as exposed based on the two

metrics, to 1, if the two metrics classify an identical set of counties as exposed to the tropical cyclone. We calculated these values for all study storms that affected ≥100 study counties (based on at least one exposure metric) during the years when all exposure data were available (1996–2011).

Data and Code Access

We have posted most study data as an open-source R package (Anderson et al. 2020a). By sharing this data as an R package, it is also accessible through cloud-based computing platforms that incorporate Jupyter notebooks, including the National Science Foundation's DesignSafe platform for natural hazards engineering research (Rathje et al. 2017). We have posted remaining data and code at https://github.com/geanders/county_hurricane_exposure.

Results

Exposure Assessment Data, Data Validation, and Software

We created and published this tropical cyclone exposure data as an R package (Anderson et al. 2020a). The package size exceeded the recommended maximum size for the Comprehensive R Archive Network (CRAN), the standard repository for publishing R packages. Therefore, we used the *drat* framework to set up our own repository to host the package (Anderson and Eddelbuettel 2017). These data include continuous county-level measurements for the closest distance of each storm, cumulative rainfall, and peak sustained surface wind. These continuous metrics can be used for classifying counties as exposed or unexposed—using thresholds selected by the user—or can be extracted directly as continuous metrics. The data set also includes binary data on flood and tornado events associated with each storm in each county. To accompany these data, we published an additional R package with software tools to explore and map the data and to integrate it with human health data sets (Anderson et al. 2020b).

We explored potential limitations in these data by comparing them with data from other available sources. For estimates of storm-associated rainfall, we compared data in the open-source package with ground-based observations in nine sample counties (Figure 2). Within these counties, storm-related rainfall measurements were well-correlated between the two data sources, with rank correlations (shown at the bottom right of each graph in Figure 2) between 0.87 and 0.98. There was some evidence that our primary rainfall metric may tend to underestimate rainfall totals in storms with extremely high rainfall, based on a few heavy-rainfall storms in Harris County, Texas; Mobile County, Alabama; Charleston County, South Carolina; and Wake County, North Carolina (Figure 2). Further, in some counties, the correlation was substantially lower when considering only tropical cyclones with cumulative local rainfall of ≥75 mm (Table S2). However, it was rare for a storm to be classified differently (under the precipitation threshold of 75 mm we used for binary exposure classification for later analysis) based on the source of precipitation data. Horizontal and vertical lines in each small plot in Figure 2 show the threshold of 75 mm, so storms in the lower left and upper right quadrants would be classified the same ("exposed" or "unexposed"), regardless of the precipitation data source, whereas storms in the upper left and lower right quadrants would be classified differently. Such cases were rare.

For peak sustained surface wind estimates, we found that the primary wind exposure metric in the open-source data generally agreed well with the wind radii reported in HURDAT2. For most storms, >90% of counties were assigned the same category of wind speed (<34 knots; 34–49.9 knots; 50–63.9 knots; ≥64 knots) by both data sources (Figure 3). Disagreement was limited to a few storms (e.g., Hurricanes Sandy in 2012 and Ike in 2008). For these two storms, the modeled wind speed in the open-source data somewhat overestimated the severity of the storms' winds at landfall but then underestimated, particularly for Hurricane Sandy, how far 34–49.9 knot winds extended from the storms' central tracks further inland (Figure S2). For epidemiology researchers who would like to conduct sensitivity analysis using both sets of wind data, we have included estimates from the HURDAT2 wind radii as a secondary measure of county-level peak sustained wind in the open-source data set (Anderson et al. 2020a).

For flood data, we compared the flood status values included in the open-source data to stream-flow measurements at USGS gages within nine sample counties (Figure 4). Across the sampled counties, streamgage data generally indicated higher discharge during periods identified as flood events based on the NOAA Storm Events database (Figure 4). There were some cases, however, where the two flooding data sources were somewhat inconsistent. For example, there were one or two tropical cyclones in several of the counties (Mobile County, Alabama; Escambia County, Florida; Fairfield County, Connecticut; and Fulton County, Georgia) with associated flood event listings but for which the total discharge across county streamflow gages was relatively low. For storms without a flood listing for the county, in most cases the total streamflow discharge in the county was relatively low, and in all but two cases, all streamgage flows were below the flooding threshold. The exceptions were for Hurricane Ida in 2009 in Fulton County, GA, and Hurricane Isaac in 2012 in Mobile County, AL.

Patterns in Tropical Cyclone Exposures

We used this storm exposure data to classify counties as exposed or unexposed to four different hazards for each tropical cyclone and then explored patterns over years with available data. Table 2 provides summary statistics describing the extent of counties assessed as exposed based on specific storm hazards to help epidemiologists in planning studies, including understanding the potential statistical power to study specific exposures, both for multistorm and single-storm studies. Across the four storm hazards considered, there was wide variation in the average number of county exposures per year (Table 2). For tropical cyclone tornadoes, there were on average approximately 40 county exposures per year in our study. County exposures were more frequent for tropical cyclone wind exposures (>160/y on average), even more frequent for tropical cyclone flood exposures (>190/y on average), and most frequent for tropical cyclone rain exposure (>290/y on average). For every hazard except tornadoes, we identified at least one tropical cyclone that exposed more than 250 counties (Table 2). However, the largest-extent tropical cyclone varied across hazards: Frances in 2004 exposed the most counties based on rain, Michael in 2018 based on wind, and Ivan in 2004 based on flooding and tornadoes (Table 2).

When we calculated and mapped the average number of exposures per decade in each county for single-hazard exposures (Figure 5), strong geographical patterns were clear. Peak

Table 2. Summary statistics for the number of county tropical cyclone exposures under each metric, with exposure assessment based on definitions in Table 1.

Metric (y available)	Mean of county exposures per year (interquartile range)	Tropical cyclone with most counties exposed (# exposed counties)
Rain (1988–2011)	298 (48, 441)	Frances, 2004 (464)
Wind (1988–2018)	162 (55, 258)	Michael, 2018 (260)
Flood (1996–2018)	197 (76, 241)	Ivan, 2004 (317)
Tornado (1988–2018)	38 (8, 42)	Ivan, 2004 (91)

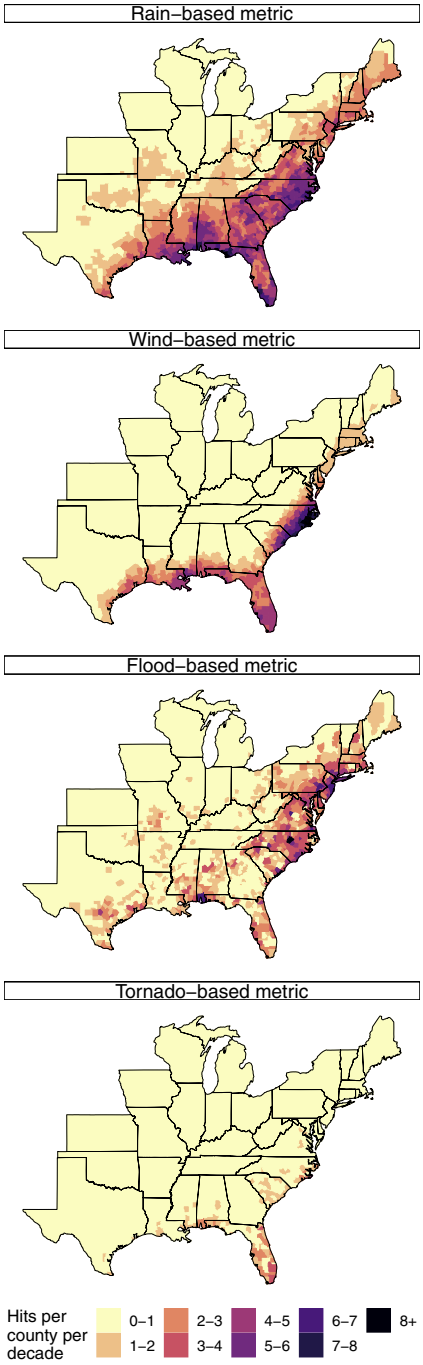

Figure 5. Average number of county-level storm exposures per decade for binary classifications based on each single-hazard exposure metric (Table 1). The years used to estimate these averages are based on years of available exposure data (rain: 1988–2011; wind: 1988–2018; flood events: 1996–2018; and tornado events: 1988–2018).

sustained wind exposure had a strong coastal pattern, with almost all exposures in counties within about 200 km (124 mi) of the coastline. Although tropical cyclone rain exposure was also more frequent in coastal areas in comparison with that in inland areas, there were also inland rain exposures in counties that were rarely or never classified as exposed based on wind. Flood-based exposures were frequently in the mid-Atlantic region, with a pattern that skewed north compared with other exposures. Rain and, to some extent, flood exposures were characterized by a pattern defined by the Appalachian Mountains, with fewer exposures west of the mountain range than to the east. Almost all tornado-based exposures were in coastal states, with many in Florida and almost none north of Maryland. Patterns were similar when analysis was restricted to years with exposure data for all four hazards available (1996–2011; Figure S3).

Agreement across Exposure Metrics

Finally, we assessed within-storm agreement between exposure classifications for each pair of hazards, as well as with a proxy measurement based on distance from the storm's track. We found that these exposure classifications typically did not agree strongly between pairs of metrics; the set of counties identified as "exposed" based on one metric often overlapped little with the set identified as "exposed" by another metric, as storms frequently brought different hazards to different locations.

Figure 6 shows as an example Hurricane Ivan in 2004. For the distance-based metric, the counties assessed as exposed follow the tropical cyclone's track. For the wind-based metric, only counties near the tropical cyclone's first landfall were assessed as exposed. For rain- and flood-based metrics, however, exposure extended to the left of the track, including counties as far north as New York and Connecticut, whereas for the tornado metric, exposed counties tended to be to the right of the track and included several counties in central North Carolina, South Carolina, and Georgia that were not identified as exposed to Ivan based on any other metric. Figure S4 provides similar maps for three other example tropical cyclones (selected because they exposed many U.S. counties based on at least one metric).

We drew similar conclusions when we investigated all 46 tropical cyclones between 1996 and 2011 (when data for all five metrics were available) for which 100 or more counties were exposed based on at least one metric (Figure 7; for the most extensive of these, storms for which 200 or more counties were exposed based on at least one metric, Tables S3–S6 provide underlying numbers comparing exposure assessments between the distance-based proxy and each hazard-based metric). In Figure 7, each row provides results for one tropical cyclone, and each box in that row shows the Jaccard coefficient for a pair of metrics. For all pairs of metrics, agreement in exposure assessment was, at best, moderate for all but a few storms. When comparing distance- and wind-based exposure assessment, only about 10% of storms had Jaccard indices higher than 0.6 (i.e., out of the counties assessed as exposed by at least one of the two metrics in the pair, 60% or more were assessed the same under both metrics). For comparisons of assessments based on other combinations of distance-, wind-, rain-, and flood-based metrics, fewer than 5% of storms had Jaccard indices above 0.6. The tornado-based metric had universally poor agreement with other metrics in county-level classification across the tropical cyclones considered.

There were a few exceptions; tropical cyclones in which exposure assessment agreed well across several of the metrics considered. For Floyd in 1999 (Figure S4) and Irene in 2011, for example, county-level classification agreed moderately to well for all pairs of exposure metrics except those involving the tornado-based metric (Figure 7). For another set of tropical cyclones (e.g., Ernesto in 2006, Bertha in 1996, and Isabel in 2003), there was moderate to good agreement for pairwise combinations of distance, rain, and wind but poor agreement for other combinations of metrics, whereas for another set of storms (e.g., Matthew in 2004 and Katrina in 2005), there was moderate to good agreement between distance and rain.

Discussion

Epidemiologic studies can help characterize which health risks are elevated during disasters, to what degree, and for whom (Ibrahim

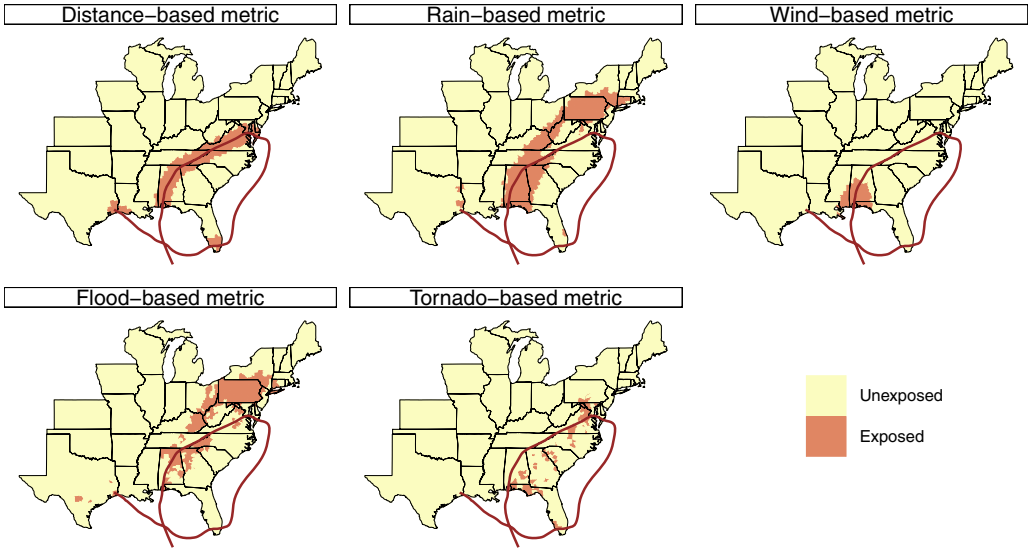

Figure 6. Counties classified as exposed to Hurricane Ivan in 2004 under each exposure metric considered (Table 1). The red line shows the track of Hurricane Ivan based on HURDAT2. Similar maps for other large-extent storms are given in Figure S4.

2005; Noji 2005). As a result, these studies help improve disaster preparedness and response (Noji 2005). However, tropical cyclones are multihazard events, making it complicated to measure exposure and so to conduct multicommunity, multiyear studies leveraging large administrative data sets. Here, we provide open-source county-level data for several tropical cyclone exposures and explore limitations in that data. Further, we explore patterns in storm exposure classifications based on different metrics, and we find that county-level tropical cyclone exposure assessments vary substantially when using different metrics. Our results can inform exposure assessment for future county-level studies of the health risk and impacts associated with tropical cyclones exposure as well as provide insights to inform epidemiological study design.

Exposure Assessment Data and Software

The open-source data generally correspond well with data from other potential sources (Figures 2–4), but there are caveats. The rainfall data are generally well-correlated with ground-based observations but may sometimes underestimate very high rainfall values (Figure 2), and in some counties the correlation was substantially lower when considering only tropical cyclones with cumulative local rainfall of $\geq$75 mm (Table S2). When rainfall data are used to create binary exposure classifications, this disagreement is unlikely to influence results because both data sources agree in identifying these as storms with high rainfall, but this disagreement would be important to consider for cases that include rainfall as a continuous measurement.

The peak sustained wind estimates are based on modeled, rather than observed, values, and although the modeled wind data generally agree well (Figure 3) with postanalysis maximum wind radii (Landsea and Franklin 2013), there were a few storms with some discrepancies. These storms (e.g., Hurricane Sandy in 2012 and Hurricane Ike in 2008) were unusually large systems for which high winds persisted well inland from landfall (Figures 3 and S2).

For the flooding data, we found that flood event status as determined based on the NOAA Storm Events listings typically agreed with measurements from USGS streamgages, with a flood event more likely to be listed if a storm elevated streamflow at streamgages across the county (Figure 4). However, there are differences between the two flooding data sets, and these differences highlight both the difficulty of measuring flood exposure at the county level and inherent challenges in using data from a storm event database for epidemiological exposure assessment. For example, there was one storm in Fulton County, Georgia, for which there was high streamflow but not an associated flood event listing (Hurricane Ida in 2009). This storm occurred in November 2009, following a month with historic rainfall and flooding in Georgia (Shepherd et al. 2011). In this case, the flooding associated with Ida was incorporated into an ongoing flood event listing, with a start date well before the 5-d window we used to temporally match storm-event listings with tropical cyclone tracks for our data.

This disagreement highlights the difficulty of large-scale pairing of storm tracks with storm event listings; without criteria for temporally matching event start dates to storm dates, many false positives would be captured, for which the occurrence of a storm during an ongoing event might be improperly attributed to the storm. However, distance and time restrictions like those we used in matching NOAA Storm Event listings with tropical cyclone locations and dates do cause occasional false negatives, as for Hurricane Ida in Fulton County, Georgia, where a storm contributes meaningfully to an ongoing event, but the event is not captured for the storm in the exposure data because its start date is not close in time to the storm's date.

There are further limitations for the flood and tornado data. These data came from the NOAA Storm Events database, which, although it is a widely used database of events maintained by NOAA, is based on reports and therefore may be prone to underreporting (Ashley and Ashley 2008; Curran et al. 2000), especially in sparsely populated areas (Witt et al. 1998; Ashley 2007), and to other reporting errors.

Although these aspects are important caveats for the data, we selected these data sources as among the best currently available for measuring each of these hazards consistently and comprehensively at a multicounty, multiyear scale. In addition to providing tropical cyclone exposure metrics for individual hazards, this data set and its associated software allow users not only to access measurements for single hazards, but also to create tropical cyclone exposure profiles based on multiple hazards or to craft exposure indices that combine hazard metrics (Chakraborty et al. 2005; Peduzzi et al. 2009). This functionality can be critical, because different hazards of tropical cyclones can act synergistically in causing impacts (Smith and Petley 2009).

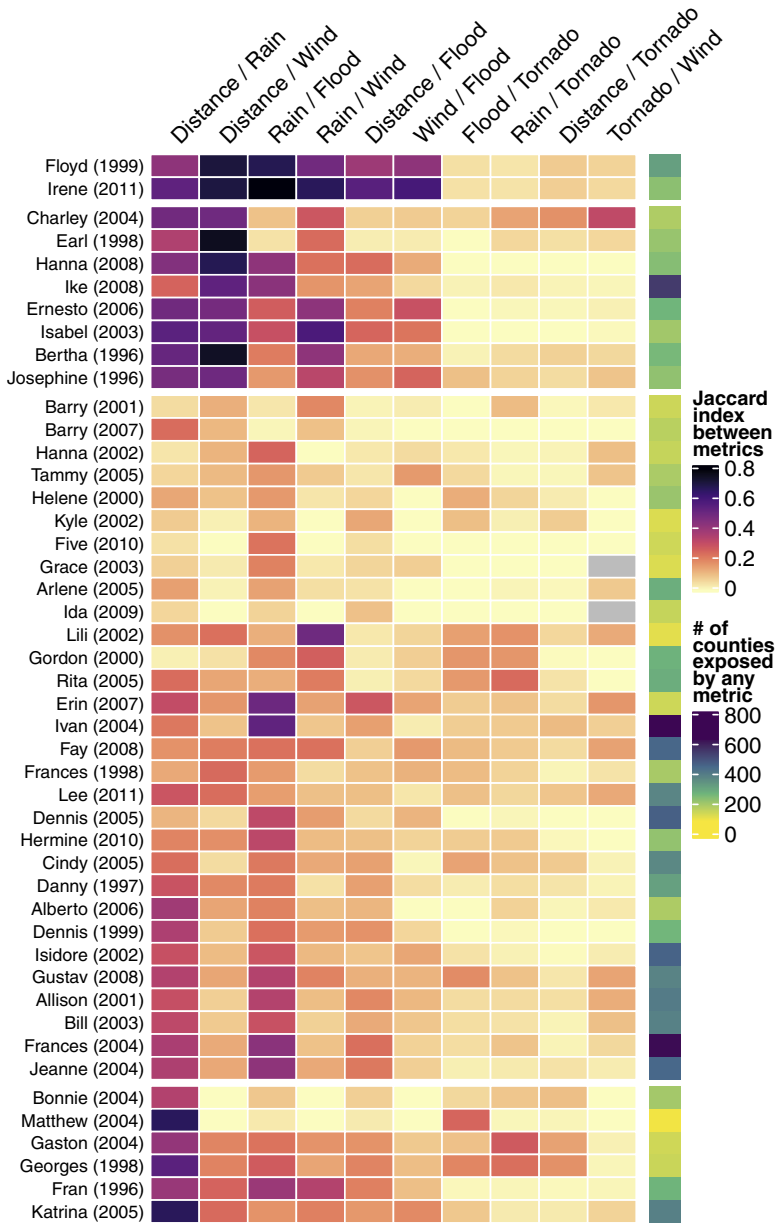

Figure 7. Agreement between exposure classifications based on different single-hazard exposure metrics for all storms between 1996 and 2011 for which at least 100 counties were exposed based on at least one metric. Each row shows one storm, and the color of each cell shows the measured Jaccard index for each pair of exposure metrics (proportion of counties classified as exposed by both metrics out of counties classified as exposed by either metric). For Grace in 2003 and Ida in 2009, there were no county exposures for either the tornado-based metric or the wind-based metric (indicated by gray squares). Colors to the right of the heatmap show the number of exposed counties based on any of the metrics, and this panel is linked with the color scale labeled "# of counties exposed by any metric." Storms are displayed within clusters that have similar patterns based on hierarchical clustering.

This data set is limited to the contiguous United States, and although expansion to global coverage would be useful and feasible, there would be some challenges. The key challenge would be for tornado and flood events. For these events, the data described in this paper drew on a U.S.-focused storm events database, and international extension of data on these hazards would require access to similar databases covering other countries. For precipitation data, whereas the reanalysis product used here (NLDAS-2) only covers the contiguous United States, other reanalysis products, as well as other types of precipitation data sets, have global coverage (Sun et al. 2018), including the National Aeronautics and Space Administration (NASA) Global Land Data Assimilation System Version 2 (GLDAS-2; Rodell et al. 2004), the Global Precipitation Measurement Mission (GPM), and the Tropical Rainfall Measuring Mission (TRMM). The wind data are generated based on a model that is currently U.S.-focused (Anderson et al. 2020c) but could be extended to other areas, although such extension would require adding new land/sea masks within the associated software, as well as accounting for differences across storm basins in wind averaging periods (Harper et al. 2010). Further, because the core of the wind model was developed based on data from Atlantic-basin storms (Willoughby et al. 2006), an extension to other areas should include separate validation and calibration to ensure it performs appropriately in those settings. Alternatively, other wind field modeling software is available that provides a global coverage, such as Geoscience Australia's Tropical Cyclone Risk Model (http://geoscienceaustralia.github.io/tcrm/). Finally, the relevant geopolitical boundaries to use for

aggregation would vary by country (e.g., municipalities in Mexico, districts in India).

The data set we present focuses on the physical hazards of a tropical cyclone. However, health impacts will often come through indirect pathways, including through damage to property and infrastructure. In future work, it would be useful to expand this data set to add data related to these pathways to allow research exploring the role of indirect pathways from tropical cyclone physical hazards to health risk. This expansion could include adding data on normalized storm-associated damages, or proxy measurements of such damages, from sources like the NOAA's Storm Events database and the U.S. Federal Emergency Management Agency's county-level disaster declarations.

Patterns in Tropical Cyclone Exposures

We found average exposures to different tropical cyclone hazards differed geographically (Figure 5). These patterns were not unexpected based on what is known about tropical cyclone hazards but still highlight variations that are critical to consider in designing studies and statistical analyses for tropical cyclone epidemiology. Further, they demonstrate the need for multihazard exposure data sets for tropical cyclone epidemiology, especially in capturing inland risks.

Tropical cyclone wind exposures had a strong coastal pattern, which is consistent with the dramatic decrease in wind intensity that typically characterizes the landfall of tropical cyclones. Tropical cyclone rain exposures tended to extend further inland in comparison with those of wind exposures up to the Appalachian mountains. This finding agrees with previous research indicating that the Appalachian mountains' topography both enhances precipitation during tropical cyclones and provides hydrological conditions for severe flooding (Sturdevant-Rees et al. 2001). Almost all tropical cyclone tornado exposures were in southern coastal states, consistent with previous evidence that tropical cyclone-related tornadoes typically occur to the right of tropical cyclone tracks in Atlantic-basin U.S. storms (Moore and Dixon 2012). It is important to note, however, that the exposure averages we calculated may be limited as estimates of long-term frequencies, because tropical cyclones follow decadal patterns (Kossin and Vimont 2007) that may not adequately captured in the available data.

Agreement between Exposure Metrics

We found that tropical cyclones tended to bring different hazards to different counties, so agreement was typically low between distance-based tropical cyclone exposure assessment and each of the single hazard-specific exposures, as well as between pairs of hazard-specific metrics (Figures 6–7 and S4). These findings align with previous results from atmospheric science and related fields on the characteristics of tropical cyclones. Although tropical cyclone rainfall and windspeed can be well-correlated when the tropical cyclone is over water (Cerveny and Newman 2000), this relationship often weakens once the hurricane has made landfall (Jiang et al. 2008). Fast-moving tropical cyclones heighten risk of dangerous winds inland (Kruk et al. 2010), whereas slow-moving tropical cyclones are likely to bring more rain (Rappaport 2000) and may cause more damage because of sustained hazardous conditions (Rezapour and Baldock 2014). Further, although the likelihood and extent of flooding during a tropical cyclone is related to the tropical cyclone's rainfall, it is also driven by factors like top soil saturation and the structure of the water basin's drainage network (Chen et al. 2015; Sturdevant-Rees et al. 2001).

Based on our results, the use of a distance-based metric to assess exposure to any of these hazards, or the use of measurements from one hazard as a proxy for exposure to any of the other hazards considered, would often introduce exposure misclassification (Figure 7, Tables S3–S6). This conclusion reinforces similar findings from a study of Florida's 2004 storm season (Grabich et al. 2015a). For some studies, such exposure misclassification might plausibly be differential. For example, tropical cyclone wind exposures tend to be concentrated in counties near the coast, whereas tropical cyclone rain exposures sometimes extend well inland. If the etiologically relevant exposure for a health outcome is extreme rainfall but exposure is classified based on measurements of wind, the probability of being misclassified as unexposed would be higher in inland counties, whereas the probability of being misclassified as exposed would be higher in coastal counties. If coastal counties differ from inland counties in either the outcome of interest or in factors associated with risk of that outcome, exposure misclassification would be differential (Savitz and Wellenius 2016). Such differential exposure misclassification could bias estimates of tropical cyclone effects either towards the null (estimating a lower or null association compared to the true association that exists) or away from the null (estimating a larger association than actually exists) (Savitz and Wellenius 2016; Armstrong 1998).

We did find a small set of tropical cyclones for which for which agreement was high across several single-hazard exposure assessments [e.g., Hurricane Floyd in 1999, Irene in 2011, Hannah in 2008, Bertha in 1996; Ernesto in 2006 (Figure 7)]. Hurricanes Floyd in 1999 and Irene in 2011 both made their first U.S. landfall in North Carolina at minor hurricane intensity (Category 2 and 1, respectively) and then skimmed the eastern coast of the United States north through New England, bringing substantial rainfall to much of the eastern coast from North Carolina north and causing extensive inland flooding in North Carolina (Floyd) and New England (Irene) (Avila and Stewart 2013; Lawrence et al. 2000). Hurricanes Hannah in 2008, Bertha in 1996, and Ernesto in 2006 also followed the eastern coastline. For these storms, the tropical cyclones' persistent proximity to water may have helped maintain wind speeds in similar patterns to rain and distance-based exposures, resulting in more similarities across exposure assessments in comparison with those of other tropical cyclones. For these storms, it may be possible to assess exposure to multiple hazards of the storm using a single metric, perhaps even a proxy like the distance between the county and the storm's track. With the data set we describe in this paper, however, there is little need to limit analysis based on exposure to a single hazard or proxy, although multihazard studies of storms with high agreement among hazard exposures should look out for modeling issues from multicollinearity. Further, for these storms it may be difficult to untangle the contribution of each hazard to the overall effect of the storm, given that several hazards have similar geographical patterns.

Limitations

The data set presented here does have several limitations, in addition to the caveats previously discussed. First, the data set is not comprehensive of all tropical cyclone hazards. For example, coastal counties can experience dangerous storm surge, which is not specifically covered in this data set (although some resulting coastal flooding is captured). We are exploring ways to include this in future versions of the data set, but to date we have focused on exposures that could affect any county, whether inland or coastal.

Second, these data are aggregated to the county level. This spatial scale allows for easy integration with health outcome data aggregated at the county level, and disaster response decisions are often made at a county or state level. Such data are often used for disaster epidemiology, because aggregated data may be easier to access than individual-level data, especially at a scale that covers many locations and years and therefore allow higher statistical

power and include a broader range of exposure levels (Wakefield and Haneuse 2008).

Such ecological exposure assessment, however, sets a common exposure level throughout the county, ignoring within-county variability, even though such variability exists. For some hazards, this within-county variation could be stark. For example, tornadoes cause very localized damage, directly along the tornado's path; a tornado can destroy homes on one side of a street but leave those on the opposite side untouched. Levels of other hazards, like storm-associated winds and rain, will also vary within a county but typically with smoother variation. In particular, it will be unlikely that a county will have one area that is exposed to extremes of these hazards but have other parts of the county that are completely unexposed, because both the wind fields and rain fields of tropical cyclones tend to be large in comparison with the size of a county.

Aggregated data can be used to infer contextual-level associations—for example, the association between county-level exposure to a storm hazard and countywide rates of a health outcome. However, ecological data are also sometimes used to infer individual-level associations (e.g., the association between personal exposure to a storm hazard and personal risk of experiencing the health outcome). Individual-level inference from ecological/aggregated data is susceptible to ecological bias (Greenland and Robins 1994; Portnov et al. 2007; Idrovo 2011). Researchers who use the data provided here for ecological studies with the aim of making individual-level inferences should be aware of this potential and could explore approaches for minimizing risk of ecological bias (e.g., Wakefield and Haneuse 2008).

Further, although many health outcome data sets are aggregated at this level, some may be aggregated at a finer spatial resolution (e.g., census tract or ZIP code) or unaggregated (i.e., point locations for each outcome). We have published the wind model used to create this data set as its own open-source R package (Anderson et al. 2020c), and it can be used to model storm-associated winds at a finer spatial resolution; however, measurements of other hazards cannot similarly be rescaled through tools we provide.

Next, we provide these data and associated software tools through R packages, and some experience in the R programming language is required to make full use of them. However, R is currently a popular programming language for environmental epidemiology, allowing the data to reach a large audience, and we are exploring options to create a web application using the Shiny platform to allow broader web-based access of the data (Chang et al. 2019).

Last, to assess patterns and agreement for binary exposure classifications, we have chosen one set of sensible thresholds for binary classifications based on continuous metrics (rainfall, maximum sustained surface wind, and distance from the storm's track), but other thresholds would be reasonable depending on hypothesized pathways for a given epidemiological study (Table S1). Results and conclusions would differ somewhat with other threshold choices. We have published code for this analysis online (https://github.com/geanders/county_hurricane_exposure), allowing other researchers to explore other threshold choices for these analyses.

Conclusions

To conduct tropical cyclone epidemiological studies that span multiple communities and storms, it is critical to have consistent and comprehensive measurements of exposure to storm hazards. Here we have created and shared a data set that provides these data for counties in the United States over multiple years. Despite some limitations in these data, they provide a powerful tool for expanding tropical cyclone epidemiology studies to more extensively leverage existing administrative health data, allowing researchers to investigate how these storms affect countywide health risk. Further, this data set provides hazard measurements that are comparable across communities and storms, allowing epidemiological researchers to design studies to explore how health risks are modified by characteristics of both the storms and the communities that are hit. The data are given in an open- source format, along with associated software tools, which allows them to be freely used and allows others to explore all associated code and to contribute additions through platforms like GitHub.

Based on our analysis shown in this paper, these data are typically in agreement with measurements from other sources of data available to characterize storm-associated hazards (e.g., ground-based monitors, streamgages, poststorm wind radii). However, researchers who are planning to use the data should explore the analyses presented in this paper to understand the strengths and weaknesses of the data. Further, our results indicate that county-level storm exposure is not well-characterized by the closest distance that a storm's central track came to a county and that exposure to one storm hazard within a county (e.g., severe winds) does not imply exposure to other hazards (e.g., excessive rainfall, flooding). As a result, it is critical that researchers consider which storm hazards are likely on the causal pathway for the outcomes they are studying and that researchers characterize storm exposure in a way that captures those specific hazards to avoid exposure misclassification.

Acknowledgments

This work was supported in part by grants from the National Institute of Environmental Health Sciences (R00ES022631), the National Science Foundation (1331399), the Department of Energy (grant no. DE-FG02-08ER64644), and a NASA Applied Sciences Program/Public Health Program Grant (NNX09AV81G). Rainfall data are based on data acquired as part of the mission of the NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences Data and Information Services Center. We have posted most study data as an open-source R package (Anderson et al. 2020a). We have posted remaining data and code at https://github.com/geanders/county_hurricane_exposure.

References

Al-Hamdan MZ, Crosson WL, Economou SA, Estes MG Jr, Estes SM, Hemmings SN, et al. 2014. Environmental public health applications using remotely sensed data. Geocarto Int 29(1):85–98, PMID: 24910505, https://doi.org/10.1080/10106049.2012.715209.

Anderson GB, Bell ML. 2011. Heat waves in the United States: mortality risk during heat waves and effect modification by heat wave characteristics in 43 US communities. Environ Health Perspect 119(2):210–218.

Anderson GB, Chen, Z. 2019. noaastormevents: Explore NOAA Storm Events Database. R Package Version 0.1.1. https://cran.r-project.org/package=noaastormevents [accessed 16 February 2019].

Anderson GB, Eddelbuettel D. 2017. Hosting data packages via drat: a case study with hurricane exposure data. R Journal 9(1):486–497. https://doi.org/10.32614/RJ-2017-026.

Anderson GB, Guo, Z. 2016. countytimezones: convert from UTC to Local Time for United States Counties. R package version 1.0.0. https://cran.r-project.org/package=countytimezones [accessed 16 February 2019].

Anderson GB, Schumacher A, Crosson W, Al-Hamdan M, Yan M, Ferreri J, Chen Z, Quiring S, Guikema S. 2020a. hurricaneexposuredata: data Characterizing Exposure to Hurricanes in United States Counties. R package version 0.1.0, https://github.com/geanders/hurricaneexposuredata [accessed 3 March 2020].

Anderson GB, Schumacher A, Guikema S, Quiring S, Ferreri J. 2020c. stormwindmodel: model Tropical Cyclone Wind Speeds. R package version 0.1.2. https://cran.r-project.org/package=stormwindmodel [accessed 3 March 2020].

Anderson GB, Yan M, Ferreri J, Crosson W, Al-Hamdan M, Schumacher A, Eddelbuettel D. 2020b. hurricaneexposure: explore and Map County-Level Hurricane Exposure in the United States. R package version 0.1.1. https://cran.r-project.org/package=hurricaneexposure [accessed 3 March 2020].

Armstrong BG. 1998. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occup Environ Med 55(10):651–656, PMID: 9930084, https://doi.org/10.1136/oem.55.10.651.

Ashley ST, Ashley WS. 2008. Flood fatalities in the United States. J Appl Meteor Climatol 47(3):805–818. https://doi.org/10.1175/2007JAMC1611.1.

Ashley WS. 2007. Spatial and temporal analysis of tornado fatalities in the United States: 1880–2005. Wea Forecasting 22(6):1214–1228, https://doi.org/10.1175/2007WAF2007004.1.

Atallah EH, Bosart LF. 2003. The extratropical transition and precipitation distribution of Hurricane Floyd (1999). Mon Wea Rev 131(6):1063–1081, https://doi.org/10.1175/1520-0493(2003)131<1063:TETAPD>2.0.CO;2.

Atallah E, Bosart LF, Aiyyer AR. 2007. Precipitation distribution associated with landfalling tropical cyclones over the eastern United States. Mon Wea Rev 135(6):2185–2206, https://doi.org/10.1175/MWR3382.1.

Avila LA, Stewart SR. 2013. Atlantic hurricane season of 2011. Mon Wea Rev 141(8):2577–2596, https://doi.org/10.1175/MWR-D-12-00230.1.

Caillouët KA, Michaels SR, Xiong X, Foppa I, Wesson DM. 2008. Increase in West Nile neuroinvasive disease after Hurricane Katrina. Emerging Infect Dis 14(5):804–808, PMID: 18439367, https://doi.org/10.3201/eid1405.071066.

Cerveny RS, Newman LE. 2000. Climatological relationships between tropical cyclones and rainfall. Mon Wea Rev 128(9):3329–3336, https://doi.org/10.1175/1520-0493(2000)128<3329:CRBTCA>2.0.CO;2.

Chakraborty J, Tobin GA, Montz BE. 2005. Population evacuation: assessing spatial variability in geophysical risk and social vulnerability to natural hazards. Nat Hazards Rev 6(1):23–33, https://doi.org/10.1061/(ASCE)1527-6988(2005)6:1(23).

Chamberlain S. 2017. rnoaa: 'NOAA' Weather Data from R. R package version 0.7.0, https://cran.r-project.org/package=rnoaa.

Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. 2019. shiny: Web Application Framework for R. R Package Version 1.3.2, https://cran.r-project.org/package=shiny.

Chen X, Kumar M, McGlynn BL. 2015. Variations in streamflow response to large hurricane-season storms in a southeastern US watershed. J Hydrometeor 16(1):55–69, https://doi.org/10.1175/JHM-D-14-0044.1.

Curran EB, Holle RL, López RE. 2000. Lightning casualties and damages in the United States from 1959 to 1994. J Climate 13(19):3448–3464, https://doi.org/10.1175/1520-0442(2000)013<3448:LCADIT>2.0.CO;2.

Currie J, Rossin-Slater M. 2013. Weathering the storm: hurricanes and birth outcomes. J Health Econ 32(3):487–503, PMID: 23500506, https://doi.org/10.1016/j.jhealeco.2013.01.004.

Czajkowski J, Kennedy E. 2010. Fatal tradeoff? Toward a better understanding of the costs of not evacuating from a hurricane in landfall counties. Popul Environ 31(1–3):121–149, https://doi.org/10.1007/s11111-009-0097-x.

Czajkowski J, Simmons K, Sutter D. 2011. An analysis of coastal and inland fatalities in landfalling US hurricanes. Nat Hazards 59(3):1513–1531, https://doi.org/10.1007/s11069-011-9849-x.

Elsberry RL. 2002. Predicting hurricane landfall precipitation: optimistic and pessimistic views from the symposium on precipitation extremes. Bull Amer Meteor Soc 83(9):1333–1339, https://doi.org/10.1175/1520-0477(2002)083<1333:PHLPOA>2.3.CO;2.

Grabich SC, Horney J, Konrad C, Lobdell DT. 2015a. Measuring the storm: methods of quantifying hurricane exposure with pregnancy outcomes. Nat Hazards Rev 17(1):06015002, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000204.

Grabich SC, Robinson WR, Engel SM, Konrad CE, Konrad C, Horney J. 2015b. County-level hurricane exposure and birth rates: application of difference-in-differences analysis for confounding control. Emerg Themes Epidemiol 12:19, PMID: 26702293, https://doi.org/10.1186/s12982-015-0042-7.

Greenland S, Robins J. 1994. Invited commentary: ecologic studies—biases, misconceptions, and counterexamples. Am J Epidemiol 139(8):747–760, PMID: 8178788, https://doi.org/10.1093/oxfordjournals.aje.a117069.

Haikerwal A, Akram M, Del Monaco A, Smith K, Sim MR, Meyer M, et al. 2015. Impact of fine particulate matter ($PM_{2.5}$) exposure during wildfires on cardiovascular health outcomes. J Am Heart Assoc 4(7):e001653, PMID: 26178402, https://doi.org/10.1161/JAHA.114.001653.

Halverson JB. 2015. Second wind: the deadly and destructive inland phase of East Coast hurricanes. Weatherwise 68(2):20–27, https://doi.org/10.1080/00431672.2015.997562.

Harper BA, Kepert JD, Ginger JD. 2010. Guidelines for converting between various wind averaging periods in tropical cyclone conditions. TD 1555. Geneva, Switzerland: World Meteorological Organization, 1–54.

Hirsch RM, De Cicco LA. 2015. User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data." In: *Techniques and Methods*. Reston, VA: U.S. Geological Survey. https://pubs.usgs.gov/tm/04/a10/ [accessed 12 March 2017].

Ibrahim MA. 2005. Unfortunate, but timely. Epidemiol Rev 27:1–2, PMID: 15958420, https://doi.org/10.1093/epirev/mxi008.

Idrovo AJ. 2011. Three criteria for ecological fallacy. Environ Health Perspect 119(8):A332, PMID: 21807589, https://doi.org/10.1289/ehp.1103768.

Jaccard P. 1901. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. Bulletin de la Société Vaudoise des. Sciences Naturelles 37:241–272, https://doi.org/10.5169/seals-266440.

Jaccard P. 1908. Nouvelles recherches sur la distribution florale. Bulletin de la Société Vaudoise des. Sciences Naturelles 44:223–270, https://doi.org/10.5169/seals-268384.

Jarvinen BR, Caso EL. 1978. NOAA Technical Memorandum NWS NHC: A tropical cyclone data tape for the North Atlantic Basin, 1886–1977: contents, limitations, and uses. National Hurricane Center. https://www.nhc.noaa.gov/pdf/NWS-NHC-1978-6.pdf [accessed 23 October 2019].

Jiang H, Halverson JB, Simpson J, Zipser EJ. 2008. Hurricane 'rainfall potential' derived from satellite observations aids overland rainfall prediction. J Appl Meteor Climatol 47(4):944–959, https://doi.org/10.1175/2007JAMC1619.1.

Kim H, Schwartz RM, Hirsch J, Silverman R, Liu B, Taioli E. 2016. Effect of Hurricane Sandy on Long Island emergency department visits. Disaster Med Public Health Prep 10(3):344–350, PMID: 26833178, https://doi.org/10.1017/dmp.2015.189.

Kinney DK, Miller AM, Crowley DJ, Huang E, Gerber E. 2008. Autism prevalence following prenatal exposure to hurricanes and tropical storms in Louisiana. J Autism Dev Disord 38(3):481–488, PMID: 17619130, https://doi.org/10.1007/s10803-007-0414-0.

Knaff JA, Slocum CJ, Musgrave KD, Sampson CR, Strahl BR. 2016. Using routinely available information to estimate tropical cyclone wind structure. Mon Wea Rev 144(4):1233–1247, https://doi.org/10.1175/MWR-D-15-0267.1.

Kossin JP, Vimont DJ. 2007. A more general framework for understanding Atlantic hurricane variability and trends. Bull Amer Meteor Soc 88(11):1767–1781, https://doi.org/10.1175/BAMS-88-11-1767.

Kruk MC, Gibney EJ, Levinson DH, Squires M. 2010. A climatology of inland winds from tropical cyclones for the eastern United States. J Appl Meteor Climatol 49(7):1538–1547, https://doi.org/10.1175/2010JAMC2389.1.

Lammars R, Anderson GB. 2017. Countyfloods: Quantify United States County-level Flood Measurements. R package version 0.0.2, https://CRAN.R-project.org/package=countyfloods.

Landsea CW, Franklin JL. 2013. Atlantic hurricane database uncertainty and presentation of a new database format. Mon Wea Rev 141(10):3576–3592, https://doi.org/10.1175/MWR-D-12-00254.1.

Lawrence M, Avila L, Beven J, Franklin J, Guiney J, Pasch R. 2000. Atlantic hurricanes: Floyd's fury highlighted another above-average year. Weatherwise 53(2):39–43, https://doi.org/10.1080/00431670009605849.

Liu JC, Wilson A, Mickley LJ, Dominici F, Ebisu K, Wang Y, et al. 2017. Wildfire-specific fine particulate matter and risk of hospital admissions in urban and rural counties. Epidemiology 28(1):77–85, PMID: 27648592, https://doi.org/10.1097/EDE.0000000000000556.

McKinney N, Houser C, Meyer-Arendt K. 2011. Direct and indirect mortality in Florida during the 2004 hurricane season. Int J Biometeorol 55(4):533–546, PMID: 20924612, https://doi.org/10.1007/s00484-010-0370-9.

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG. 2012. An overview of the Global Historical Climatology Network—Daily database. J Atmos Oceanic Technol 29(7):897–910, https://doi.org/10.1175/JTECH-D-11-00103.1.

Milojevic A, Armstrong B, Wilkinson P. 2017. Mental health impacts of flooding: a controlled interrupted time series analysis of prescribing data in England. J Epidemiol Community Health 71(10):970–973, PMID: 28860201, https://doi.org/10.1136/jech-2017-208899.

Mongin SJ, Baron SL, Schwartz RM, Liu B, Taioli E, Kim H. 2017. Measuring the impact of disasters using publicly available data: application to 2012 Hurricane Sandy. Am J Epidemiol 186(11):1290–1299, PMID: 29206990, https://doi.org/10.1093/aje/kwx194.

Moore TW, Dixon RW. 2012. Tropical cyclone–tornado casualties. Nat Hazards 61(2):621–634, https://doi.org/10.1007/s11069-011-0050-z.

NOAA (National Oceanic and Atmospheric Administration)/National Weather Service. 2020. NOAA Storm Prediction Center WCM page. https://www.spc.noaa.gov/wcm [accessed 1 September 2020].

NOAA NCEI (NOAA National Centers for Environmental Information). 2020. Storm Events Database, https://www.ncdc.noaa.gov/stormevents/.

Noji EK. 2005. Disasters: introduction and state of the art. Epidemiol Rev 27(1):3–8, PMID: 15958421, https://doi.org/10.1093/epirev/mxi007.

Peduzzi P, Dao H, Herold C, Mouton F. 2009. Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index. Nat Hazards Earth Syst Sci 9(4):1149–1159, https://doi.org/10.5194/nhess-9-1149-2009.

Phadke AC, Martino CD, Cheung KF, Houston SH. 2003. Modeling of tropical cyclone winds and waves for emergency management. Ocean Engineering 30(4):553–578, https://doi.org/10.1016/S0029-8018(02)00033-1.

Portnov BA, Dubnov J, Barchana M. 2007. On ecological fallacy, assessment errors stemming from misguided variable selection, and the effect of aggregation on the outcome of epidemiological study. J Expo Sci Environ Epidemiol 17(1):106–121, PMID: 17033679, https://doi.org/10.1038/sj.jes.7500533.

Rappaport EN, Blanchard BW. 2016. Fatalities in the United States indirectly associated with Atlantic tropical cyclones. Bull Amer Meteor Soc 97(7):1139–1148, https://doi.org/10.1175/BAMS-D-15-00042.1.

Rappaport EN. 2000. Loss of life in the United States associated with recent Atlantic tropical cyclones. Bull Amer Meteor Soc 81(9):2065–2073, https://doi.org/10.1175/1520-0477(2000)081<2065:LOLITU>2.3.CO;2.

Rappaport EN. 2014. Fatalities in the United States from Atlantic tropical cyclones: new data and interpretation. Bull Amer Meteor Soc 95(3):341–346, https://doi.org/10.1175/BAMS-D-12-00074.1.

Rathje EM, Dawson C, Padgett JE, Pinelli J-P, Stanzione D, Adair A, et al. 2017. DesignSafe: new cyberinfrastructure for natural hazards engineering. Nat Hazards Rev 18(3):06017001, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000246.

Rezapour M, Baldock TE. 2014. Classification of hurricane hazards: the importance of rainfall. Wea Forecasting 29(6):1319–1331, https://doi.org/10.1175/WAF-D-14-00014.1.

Rocha LA, Fromknecht CQ, Redman SD, Brady JE, Hodge SE, Noe RS. 2017. Medicolegal death scene investigations after natural disaster- and weather-related events: a review of the literature. Acad Forensic Pathol 7(2):221–239, PMID: 28845205, https://doi.org/10.23907/2017.023.

Rodell M, Houser PR, Jambor UEA, Gottschalck J, Mitchell K, Meng C-J, et al. 2004. The Global Land Data Assimilation System. Bull Amer Meteor Soc 85(3):381–394, https://doi.org/10.1175/BAMS-85-3-381.

Rui H, Mocko D. 2014. README Document for North America Land Data Assimilation System Phase 2 (NLDAS-2) Products. Goddard Earth Sciences Data and Information Services Center. https://disc.gsfc.nasa.gov/datasets/NLDAS_NOAH0125_MC_002/summary [accessed 23 October 2019].

Santos-Burgoa C, Sandberg J, Suárez E, Goldman-Hawes A, Zeger S, Garcia-Mexa A, et al. 2018. Differential and persistent risk of excess mortality from Hurricane Maria in Puerto Rico: a time-series analysis. Lancet Planet Health 2(11):e478–e488, PMID: 30318387, https://doi.org/10.1016/S2542-5196(18)30209-2.

Santos-Lozada AR, Howard JT. 2018. Use of death counts from vital statistics to calculate excess deaths in Puerto Rico following Hurricane Maria. JAMA 320(14):1491–1493, PMID: 30073274, https://doi.org/10.1001/jama.2018.10929.

Savitz DA, Wellenius, GA. 2016. *Interpreting Epidemiologic Evidence: Connecting Research to Applications*. New York: Oxford University Press.

Severson R, Anderson GB. 2016. Countyweather: Compiles Meteorological Data for U.S. Counties. R Package Version 0.1.0, https://CRAN.R-project.org/package=countyweather.

Shepherd M, Mote T, Dowd J, Roden M, Knox P, McCutcheon SC, et al. 2011. An overview of synoptic and mesoscale factors contributing to the disastrous Atlanta flood of 2009. Bull Amer Meteor Soc 92(7):861–870, https://doi.org/10.1175/2010BAMS3003.1.

Smith K, Petley DN. 2009. *Environmental Hazards: Assessing Risk and Reducing Disaster*. New York: Taylor and Francis.

Son J-Y, Lee J-T, Anderson GB, Bell ML. 2012. The impact of heat waves on mortality in seven major cities in Korea. Environ Health Perspect 120(4):566–571, PMID: 22266672, https://doi.org/10.1289/ehp.1103759.

Spearman C. 1904. The proof and measurement of association between two things. By C. Spearman, 1904. Am J Psychol 15(1):72–101, PMID: 3322052, https://doi.org/10.2307/1412159.

Sturdevant-Rees P, Smith JA, Morrison J, Baeck ML. 2001. Tropical storms and the flood hydrology of the central Appalachians. Water Resour Res 37(8):2143–2168, https://doi.org/10.1029/2000WR900310.

Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu K-L. 2018. A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev Geophys 56(1):79–107, https://doi.org/10.1002/2017RG000574.

Swerdel JN, Janevic TM, Cosgrove NM, Kostis JB, Myocardial Infarction Data Acquisition System (MIDAS 24) Study Group. 2014. The effect of Hurricane Sandy on cardiovascular events in New Jersey. J Am Heart Assoc 3(6):e001354, PMID: 25488295, https://doi.org/10.1161/JAHA.114.001354.

U.S. Census Bureau. 2020. Centers of Population. https://www.census.gov/geographies/reference-files/2010/geo/2010-centers-population.html [accessed 1 September 2020].

U.S. CDC (U.S. Centers for Disease Control). 2019. North America Land Data Assimilation System (NLDAS) Daily Precipitation years 1979–2011 on CDC WONDER Online Database, http://wonder.cdc.gov/wonder/help/Precipitation.html.

USGS (U.S. Geological Survey). 2020. National Weather Information System data available on the world wide web (USGS Water Data for the Nation), https://waterdata.usgs.gov/nwis.

Wakefield J, Haneuse SJ-P. 2008. Overcoming ecologic bias using the two-phase study design. Am J Epidemiol 167(8):908–916, PMID: 18270370, https://doi.org/10.1093/aje/kwm386.

Willoughby HE, Darling RWR, Rahn ME. 2006. Parametric representation of the primary hurricane vortex. Part II: a new family of sectionally continuous profiles. Mon Wea Rev 134(4):1102–1120, https://doi.org/10.1175/MWR3106.1.

Witt A, Eilts MD, Stumpf GJ, Johnson JT, Mitchell EDW, Thomas KW. 1998. An enhanced hail detection algorithm for the WSR-88D. Wea Forecasting 13(2):286–303, https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2.

Zandbergen PA. 2009. Exposure of US counties to Atlantic tropical storms and hurricanes, 1851–2003. Nat Hazards 48(1):83–99, https://doi.org/10.1007/s11069-008-9250-6.

Zhu L, Quiring SM. 2013. Variations in tropical cyclone precipitation in Texas (1950 to 2009). J Geophys Res Atmos 118(8):3085–3096, https://doi.org/10.1029/2012JD018554.