Dogs That Shouldn't Bark

Exploring Evidence of Residual Confounding in Tropical Cyclone Epidemiology Using Negative Controls

Brooke Anderson and Meilin Yan
Colorado State University, Peking University

: brooke.anderson@colostate.edu
: @gbwanderson
: github.com/geanders

The Adventure of Silver Blaze

Source: Doyle, The Adventure of Silver Blaze, 1892

The dog that didn't bark

Gregory (Scotland Yard detective): "Is there any other point to which you would wish to draw my attention?"

Sherlock Holmes: "To the curious incident of the dog in the night-time."

Gregory: "The dog did nothing in the night-time."

Sherlock Holmes: "That was the curious incident."

Source: Doyle, The Adventure of Silver Blaze, 1892

The dog that didn't bark

Sherlock Holmes: "I had grasped the significance of the silence of the dog . . . Obviously the midnight visitor was someone whom the dog knew well."

Source: Doyle, The Adventure of Silver Blaze, 1892

Negative controls in observational studies

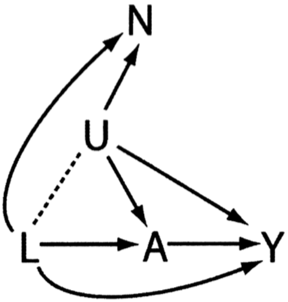


FIGURE 2. Causal diagram showing an ideal negative control outcome N for use in evaluating studies of the causal relationship between exposure A and outcome Y. N should ideally have the same incoming arrows as Y, except that A does not cause N; to the extent this criterion is met, N is called U-comparable to Y.

Source: Lipsitch et al., Epidemiology, 2010

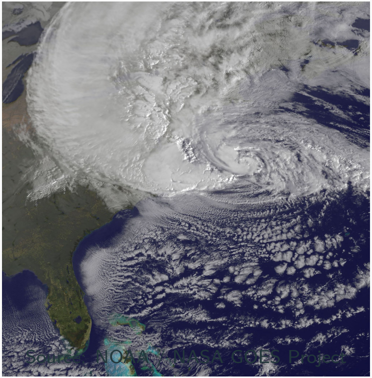
Negative controls

Negative controls have a long tradition in environmental epidemiology.

In studies that aimed to estimate the relationship between particulate matter and respiratory outcomes, examples include:

- Substituting exposure with particulate matter from the same area and day of year but a different year (Lumley and Shepard, *Environmentrics*, 2000)
- Substituting exposure with particulate matter from a different city, 140 miles away (Lumley and Shepard, *Environmentrics*, 2000)
- Substituting the outcome with accidental deaths (Borja-Aburto et al., *American Journal of Epidemiology*, 1997)
- Substituting the outcome with appendicitis (Shepard et al., *Epidemiology*, 1999)

Health risks associated with Hurricane Sandy (2012)



Health risks in storm-affected areas

- Change in patterns of emergency department visits (Kim et al. 2016)
- Increased outpatient cases of food and waterborne disease among elderly (Bloom et al. 2016)
- Increased rate of myocardial infarctions (Swerdel et al. 2014)
- Increased hospitalizations for dehydration (Lee et al. 2016)
- Difficulty obtaining medical care, medications, and medical equipment (Davidow et al. 2016)

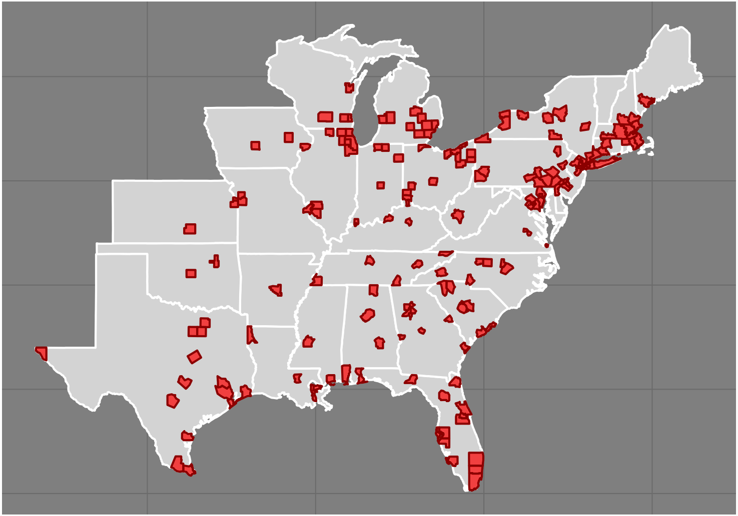
Study question

Typically, how do county-level hospitalization rates change during county-level exposure to tropical cyclone winds of ≥ 21 m/s for cardiovascular and respiratory outcomes among Medicare beneficiaries?

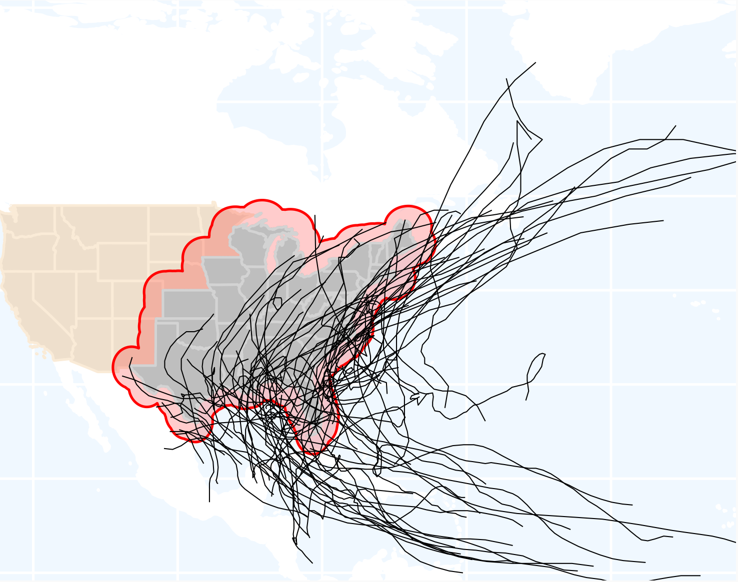
Study counties

Counties considered in our study

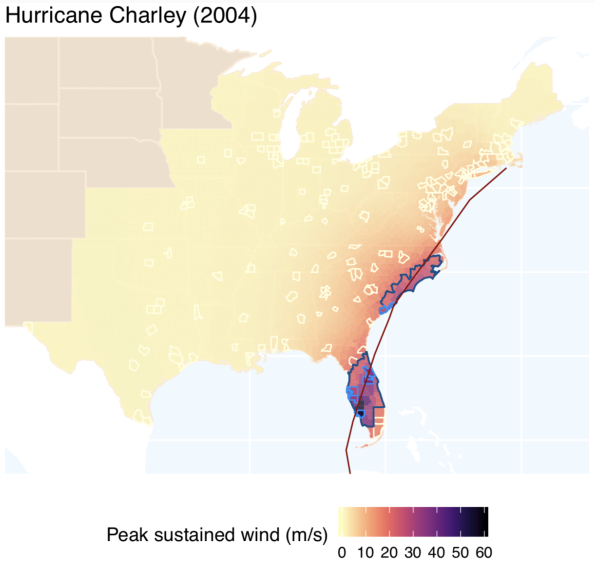
180 urban counties in the eastern US



Storms near or over the US, 1999–2010

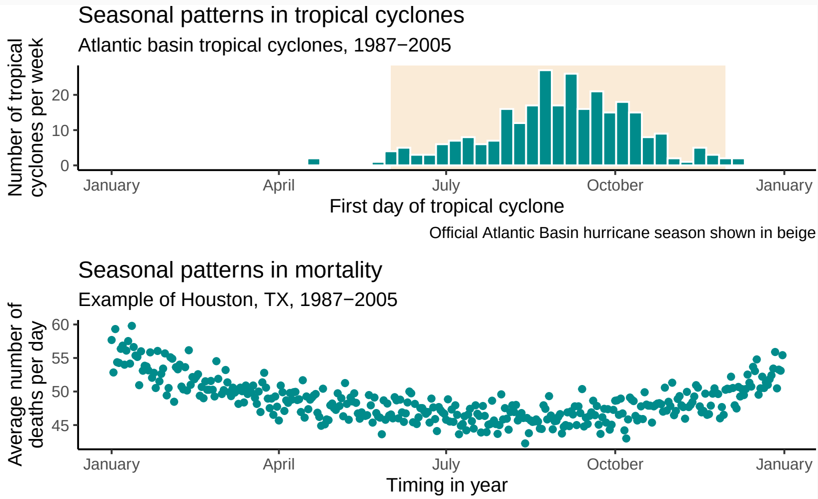


Wind field modeling and exposure assessment

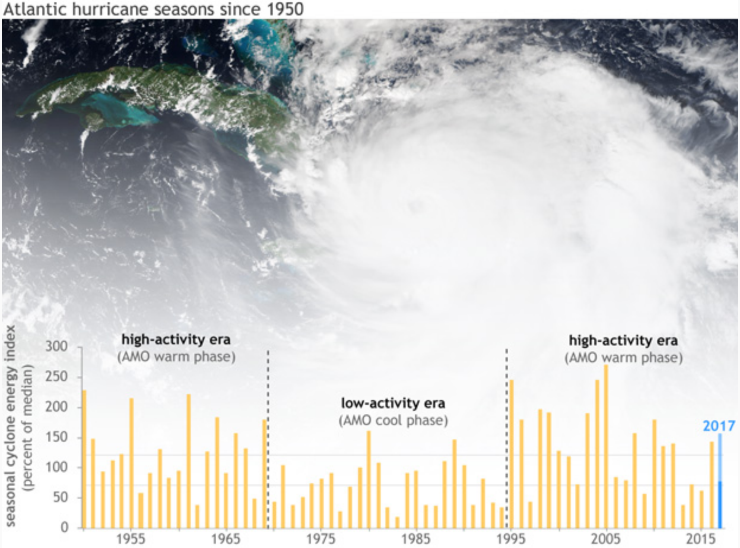


Study counties assessed as unexposed are outlined in yellow, while those assessed as exposed are outlined in light blue.

Potential for seasonal confounding



Potential for confounding by long-term trends



Source: climate.gov

Matched analysis

We matched each storm-exposed day to ten unexposed days in the county, randomly selected from candidate days that were:

1. in a different year
2. within a seven-day window of the exposure's day of year
3. outside a three-day window of a different storm-exposed day for the county
4. outside September 11–24, 2001

Estimating tropical cyclone-hospitalization associations

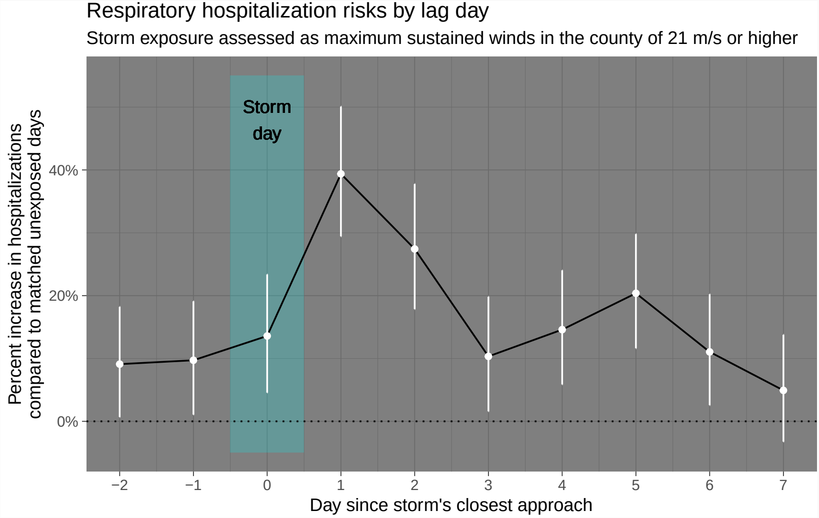
We then used this matched data to fit a generalized linear mixed-effect model of hospitalization rates in association with tropical cyclone exposure:

$$log[E(Y_t^c)] = log(n_t^c) + \alpha + \alpha_c + \sum_{l=-2}^{7} \beta_l x_{t-l}^c + \delta Z_t + \gamma D_t$$

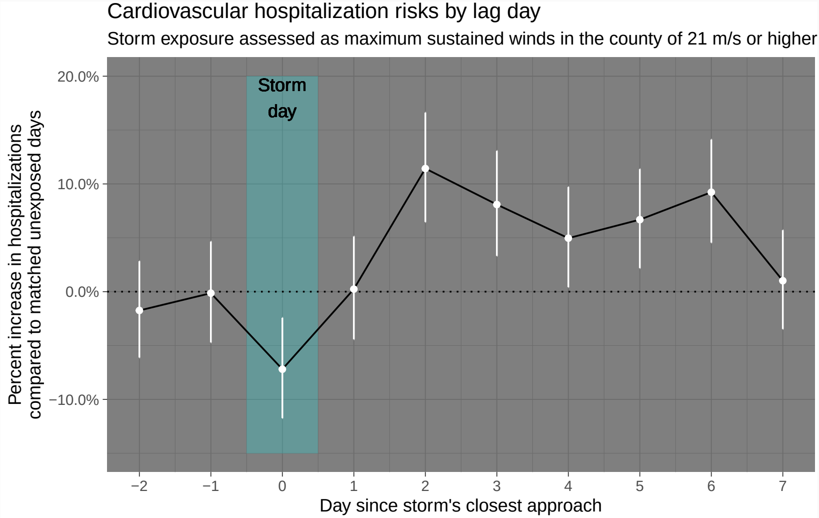
where:

- Y_t is the total count of hospital admissions on day t in community c
- n_T^c is an offset for the number of unhospitalized Medicare beneficiaries in the county on day t in community c
- α is the model intercept
- α_c is a random effect for study county
- x_{t-l} is an indicator variable for storm exposure, with associated lag-specific coefficients β_l
- Z_t is the year of day t, fit as a factor and with associated coefficient δ
- D_t is the day of week of day t, with associated coefficient γ

Hospitalization risks by lag day



Hospitalization risks by lag day



Negative control analysis

To check for residual confounding by long-term temporal trends, we conducted a negative control analysis, where we substituted the true outcome with a negative control outcome hospitalizations a bit **before** each real storm day.

Negative controls in observational studies

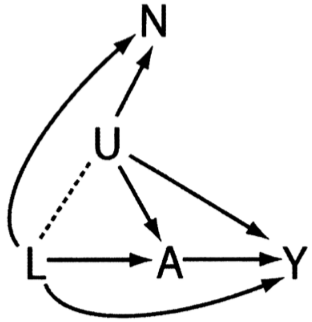
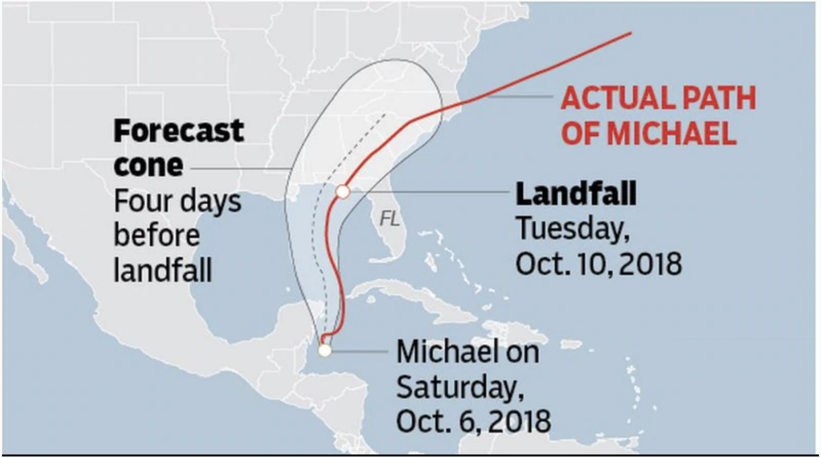


FIGURE 2. Causal diagram showing an ideal negative control outcome N for use in evaluating studies of the causal relationship between exposure A and outcome Y. N should ideally have the same incoming arrows as Y, except that A does not cause N; to the extent this criterion is met, N is called U-comparable to Y.

Source: Lipsitch et al., Epidemiology, 2010

Negative control analysis

A few days before landfall, the storm may already cause changes to the community.



Source: sun-sentinel.com

Negative control analysis

Two weeks before landfall is typically far too early to have any idea where the storm will hit or how severe it will be.

Source: weather.com

Estimating tropical cyclone-hospitalization associations

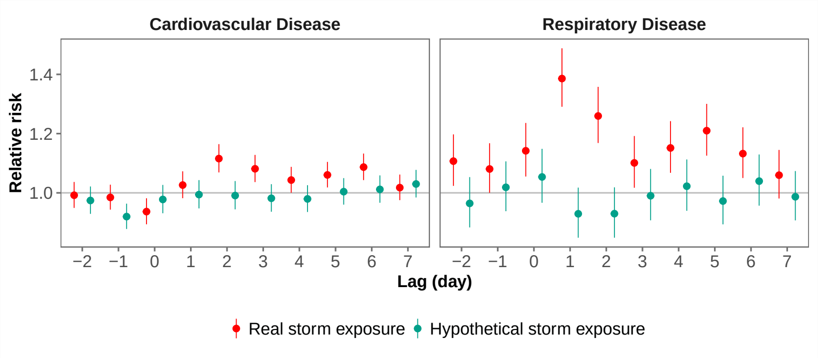
We then used this matched data to fit a generalized linear mixed-effect model of hospitalization rates in association with tropical cyclone exposure:

$$log[E(Y^c_{t-14})] = log(n^c_{t-14}) + \alpha + \alpha_c + \sum_{l=-2}^{7} \beta_l x^c_{t-l} + \delta Z_t + \gamma D_t$$

where:

- Y_{t-14} is the total count of hospital admissions on day $t-14$ in community c
- n^c_{t-14} is an offset for the number of unhospitalized Medicare beneficiaries in the county on day $t-14$ in community c
- α is the model intercept
- α_c is a random effect for study county
- x_{t-l} is an indicator variable for storm exposure, with associated lag-specific coefficients β_l
- Z_t is the year of day t, fit as a factor and with associated coefficient δ
- D_t is the day of week of day t, with associated coefficient γ

Negative control analysis—negative controls



Negative controls for calibration

- Tchetgen Tchetgen, 2013. **The control outcome calibration approach for causal inference with unobserved confounding.** *American Journal of Epidemiology.*

Leveraging temporal logic

- Flanders et al., 2017. **A new method for partial correction of residual confounding in time-series and other observational studies.** *American Journal of Epidemiology.*
- Miao and Tchetgen Tchetgen, 2017. **Bias attenuation and identification of causal effects with multiple negative controls.** *American Journal of Epidemiology.*

Negative controls in non-ideal settings

- Weisskopf et al., 2016. **On the use of imperfect negative control exposures in epidemiologic studies.** *Epidemiology.*
- Sanderson et al., 2018. **Negative control exposure studies in the presence of measurement error: implications for attempted effect estimate calibration.** *International Journal of Epidemiology.*

Acknowledgements

Colorado State University

- **Meilin Yan**
- Ander Wilson
- Joshua Ferreri
- Andrea Schumacher

NASA Marshall Space Flight Center

- Mohammad Al-Hamdan
- William Crosson

Debian / University of Illinois

- Dirk Eddelbuettel

Johns Hopkins (Public Health)

- Roger Peng

Harvard (Public Health)

- Francesca Dominici
- Yun Wang

University of Michigan

- Seth Guikema

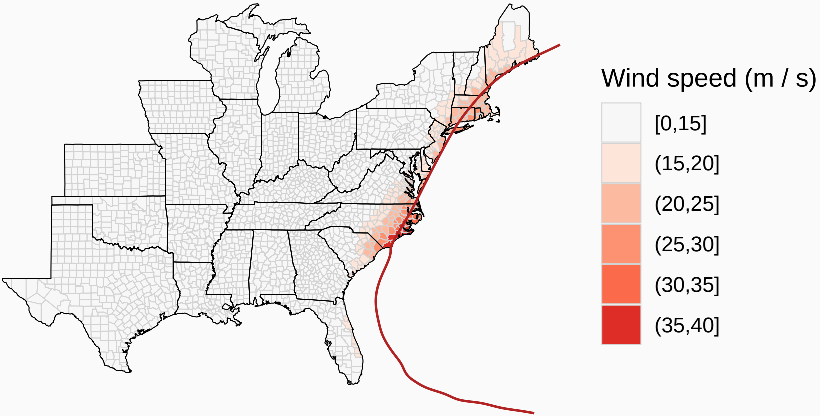
Ohio State University

- Steven Quiring

Appendix

The `hurricaneexposure` package can be used to map exposures for specific storms:

```
map_counties(storm = "Floyd-1999", metric = "wind")
```



hurricaneexposure package

‘hurricaneexposure‘ package

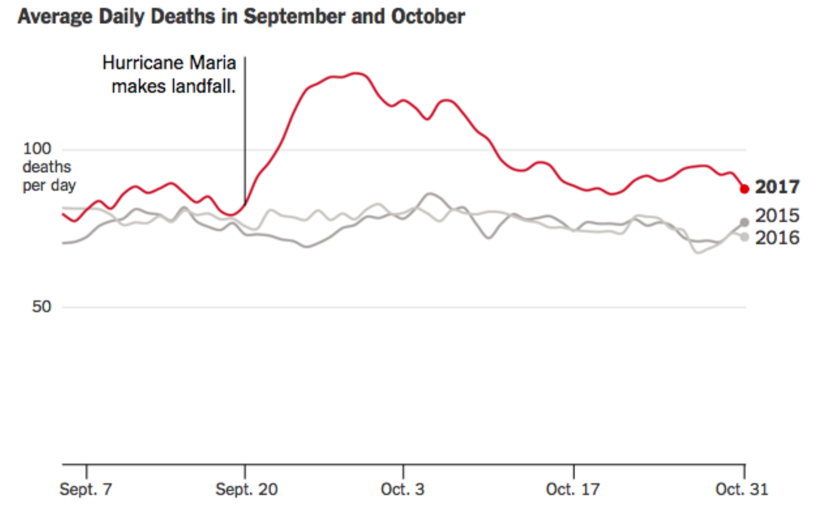
Create county-level exposure time series for tropical storms in U.S. counties. Exposure can be determined based on several hazards (e.g., wind, rain, tornadoes), with user-specified thresholds. On CRAN.

```
county_wind(counties = c("22071", "51700"), wind_limit = 21,
            start_year = 1995, end_year = 2005)
```

```
##       storm_id  fips closest_date vmax_sust
## 1  Bertha-1996 51700   1996-07-13  30.47184
## 2   Danny-1997 22071   1997-07-18  25.86265
## 3 Georges-1998 22071   1998-09-28  25.83318
## 4   Floyd-1999 51700   1999-09-16  23.47760
```

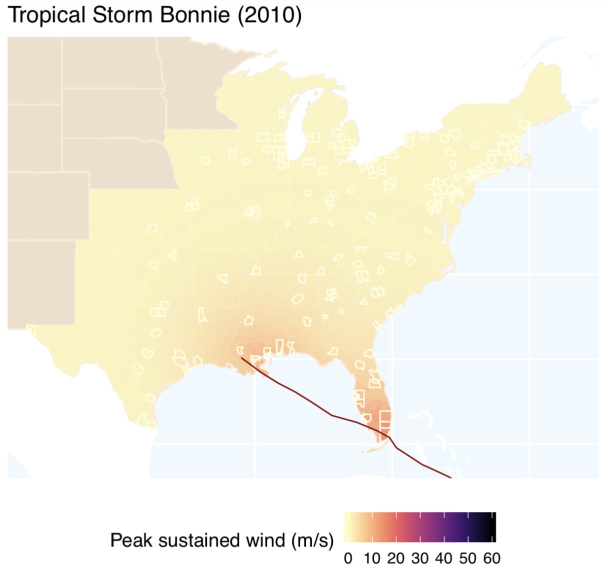
Potential for extended effects

Evidence from Hurricane Maria in Puerto Rico.



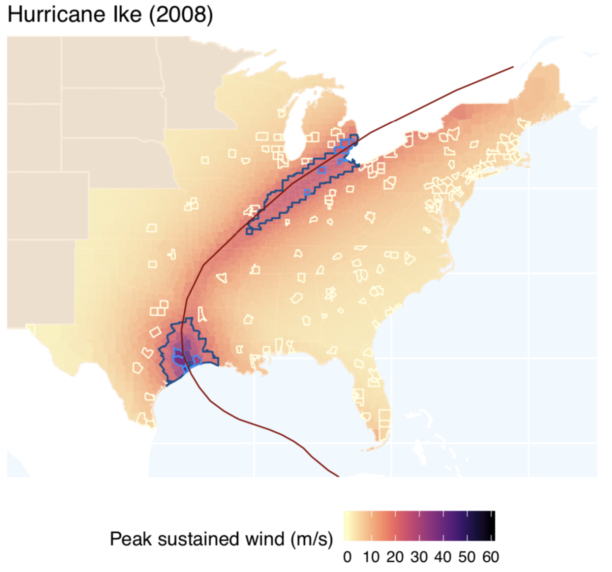
Source: The New York Times

Wind field modeling and exposure assessment



Study counties assessed as unexposed are outlined in yellow, while those assessed as exposed are outlined in light blue.

Wind field modeling and exposure assessment



Study counties assessed as unexposed are outlined in yellow, while those assessed as exposed are outlined in light blue.