**Measuring power outage exposure using a novel national dataset of power outages, 2018-2020**

**Introduction:**

Power outage incidence is increasing[[1]](#endnote-1),[[2]](#endnote-2). Climate change has increased the frequency and intensity of severe weather, the most common cause of power outages[[3]](#endnote-3),[[4]](#endnote-4),[[5]](#endnote-5). At the same time, the United States electrical grid is aging[[6]](#endnote-6),[[7]](#endnote-7). Grid components have not been modernized to withstand the previously rare extreme heat, wind, and precipitation now commonplace with climate change[[8]](#endnote-8). As a result, US electrical customers experienced an average of 8 hours without power in 2020, the longest duration on record[[9]](#endnote-9).

Power outages pose serious health risks to vulnerable people. For those who use life-sustaining electricity-dependent medical equipment such as at-home ventilators and oxygen tanks, loss of electricity can be life-threatening[[10]](#endnote-10). In children, outages increase accidents and injuries related to generator and natural gas use[[11]](#endnote-11). Power outages render air conditioners, heaters, and tap water unavailable. This can cause heat exposure, cold exposure, and dehydration in affected populations. Older adults are susceptible to stroke, myocardial infarction, and other adverse cardiorespiratory outcomes from these extreme temperature exposures and dehydration[[12]](#endnote-12),[[13]](#endnote-13),[[14]](#endnote-14),[[15]](#endnote-15). Outages also increase pediatric asthma emergencies from heat and humidity exposure absent air conditioning[[16]](#endnote-16).

Despite the health risks of power outage, data describing power outage exposure is extremely limited[[17]](#endnote-17),[[18]](#endnote-18), constraining research. Only one New York State-wide dataset describes outage exposure across space and time[[19]](#endnote-19), and most studies of power outage rely on this single dataset. The remaining studies use large-scale events such as hurricanes or disasters as a surrogate for power outage exposure[[20]](#endnote-20),[[21]](#endnote-21). These studies consider everyone in a city or county exposed to the large-scale event as exposed to power outage, in days or weeks following the event. Studies based on single events cannot disentangle the health effects of power outage exposure from disaster exposure, and cannot estimate exposure-response relationships between power outage exposure and health outcomes. Quantifying the risks of power outage exposure is essential to prevent cardiorespiratory events in older adults and electricity-dependent medical equipment users, and accidents and asthma in children.

In our previous work, we created a new national dataset of hourly power outage exposure for all counties in the continental United States[[22]](#endnote-22). This dataset will allow us to characterize exposure-response relationships between power outage and health outcomes nationally, by region, and within vulnerable populations.

However, even with these new data, major challenges with power outage exposure assessment remain when using these data for epidemiological studies of power outage exposure and health outcomes. First, there is no established strategy to measure power outage exposure in the literature[[23]](#endnote-23). A single strategy to describe power outage exposure would allow comparability and aggregation of results across studies. Second, when measuring exposure, we must specify how long a power outage must be to constitute an exposure. However, there is no literature on the duration at which a power outage begins to cause health effects, for any outcome. Incorrect assumptions about the health-relevant duration have the potential to bias the results of epidemiological studies of power outage and health outcomes. Finally, both the new national dataset and existing New York State data are missing large percentages of observations[[24]](#endnote-24),[[25]](#endnote-25), which could also cause bias in studies of outage and health outcomes.

In this paper we will address these exposure measurement issues by developing a strategy for measuring power outage exposure. Then, we will run simulations to test how assumptions about health relevant duration of outage and missingness could bias the results of an epidemiological study of the health effects of power outage. Our results will allow us and other researchers to consistently define and measure power outage exposure using the datasets currently available, while minimizing potential bias in future epidemiological studies of power outages and health outcomes.

**Methods:**

**Power outage data structure**

In our previous work, we created a national dataset of power outage exposure[[26]](#endnote-26). We purchased raw power outage data from PowerOutages.us. Most utility websites report the number of customers without power by neighbourhood or city in real time, so customers can track outages. To create this dataset, PowerOutages.us scraped counts of customers without power from utility website APIs covering the continental US, in real time, every ten minutes from 2018-2020[[27]](#endnote-27).

The resulting dataset contains hourly counts of customers without power for each continental US county from January 1st, 2018 to December 31st, 2020. Utilities define a ‘customer’ as a grid connection, which can correspond to a household, apartment building, or business[[28]](#endnote-28). Counts of customers out reported in this dataset do not necessarily track the same customers: if 10 customers are reported without power in two subsequent hours in one county, the data do not contain information about whether the same 10 households were out. The data only show that 10 households were out in each hour.

The New York State power outage dataset is structured similarly – counts of customers without power are reported by hour by power operating divisions. Power operating divisions are a geographic unit varying in size throughout the state.

**Strategy to measure power outage**

To measure daily binary power outage exposure in the PowerOutages.us dataset and New York State dataset, we propose the following strategy. This strategy could be implemented in either the New York State data or PowerOutages.us data, but here we use the PowerOutages.us dataset as an example.

Since the health-relevant duration of power outage may change depending on the health outcome being studied, we suggest a definition of power outage exposure which is flexible to identify power outages of varying lengths. Although continuous measures of power outage are possible, we chose to measure daily binary exposure because binary metrics are easily interpretable by policy makers and non-scientists. The measurement strategy we propose here is also similar to previous definitions of power outage exposure used in the literature.

To determine if a county-day was exposed to power outage, first we considered each hour alone. We considered a county-hour exposed to power outage if the percentage of customers without power in county *i* during hour *j* exceeded an arbitrary cut point – for example, 10% of the customers. If more than 10% of customers served in county *i* were without power in hour *j*, there was a power outage in county *i* for hour *j*.

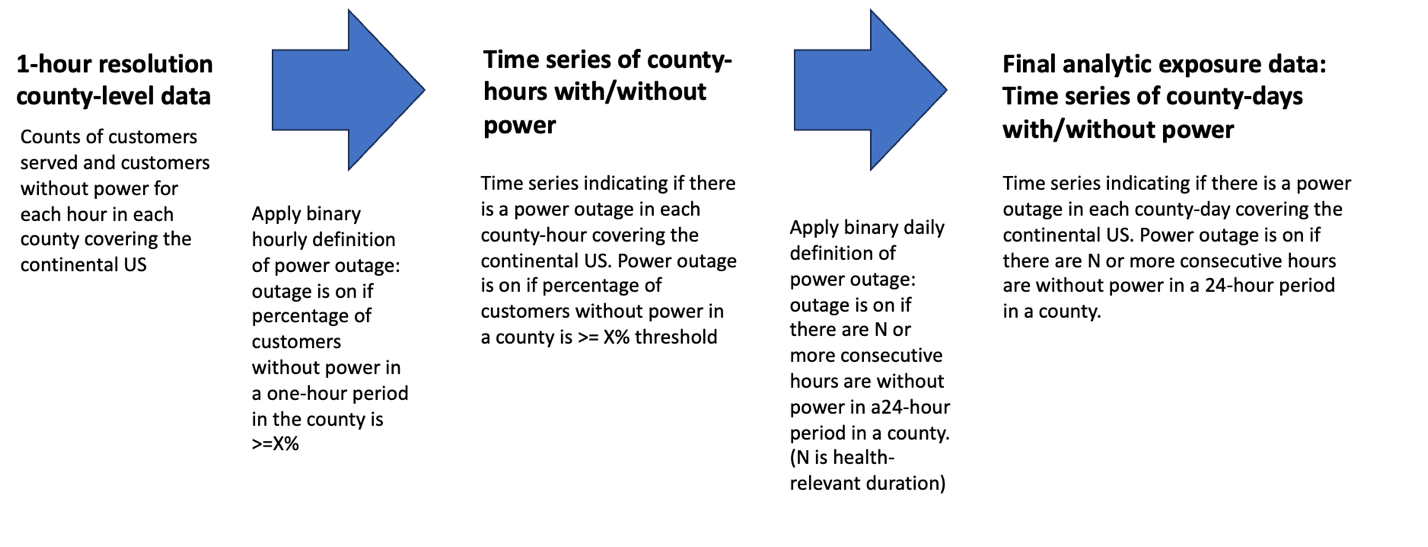
We then summarized this hourly exposure to the daily level. We chose a health-relevant duration (for example, 8 hours); this could be any duration chosen by a researcher. We considered a county-day as exposed if there were at least 8 consecutive hours of ‘power outage on’ (customers without power percentages > 10% of county) in that county on that day. Power outages could last more than 24 hours. We also considered a county-day as exposed if a power outage lasting longer than the health-relevant duration ended on that county-day.

When using this definition for 8+ hour power outages, if a certain county-day is exposed to power outage, this does not mean that there will be 10% of people in that county without power for at least 8 hours that day. Because a utility customer is a grid connection, one utility customer could be a household or apartment building. Also, counts of customers without power over time in the New York State and PowerOutages.us datasets do not track the same customers. If a county-day is exposed to power outage according to this definition, it means that many individuals were likely without power for close to 8 hours in that county on that day. This definition measures exposure for a spatial unit, rather than individuals.

Relatedly, there is exposure misclassification inherent in this definition: when the county is ‘exposed’, some customers in the county will be without electricity and others will still have electricity. Other studies of power outage exposure using a similar exposure definition have dealt with this exposure misclassification by conducting sensitivity analyses varying the cut point after which a unit is considered exposed to power outage. For example, Northrop et al. considered a spatial unit exposed to power outage if more than 10% of the customers served in that unit were without power, and conducted two sensitivity analyses where they considered a spatial unit exposed to power outage if more than 20% and 30% of the customers served in that unit were without power[[29]](#endnote-29). As the cut point percentage increases, the specificity of this definition of power outage increases.

We propose using this strategy for measuring power outage exposure, while always conducting a sensitivity analysis on the cut point. This definition of power outage exposure allows us to specify a health-relevant duration of power outage for the health outcome of interest. It allows us to compare spatial units with different populations of customers served. It is also readily interpretable by policy makers. Exposure misclassification is inherent in a daily binary definition of power outage exposure, and this exposure measurement strategy allows us to conduct a sensitivity analysis by varying the cut-point used.

Here's a draft of a flowchart that might help keep this on the rails in the future?



**Bias from misidentifying health-relevant duration and missing data**

Power outage duration matters. For some power outage-related health outcomes, there may be threshold effects where power outages longer than some duration cause adverse health outcomes, but outages shorter than that duration do not. For example, back-up batteries for many life-sustaining electricity dependent medical devices last 8 hours. 8+ hour power outages may affect the health of those using oxygen tanks and at-home ventilators, because after 8 hours of power outage, the batteries in these devices die. However, shorter outages may have no effect. Our definition of power outage exposure allows researchers to specify a power outage duration of interest to model this scenario. However, there is no literature describing how long power outages must be to cause health effects, with respect to any health outcome – researchers can only hypothesize. In the following simulation, we attempt to quantify the magnitude and direction of bias introduced when researchers assume one length of power outage (for example, 8+ hour outages) cause health outcomes, but outages of a different length (for example, either 4+ or 12+ hour outages) actually cause health effects.

Additionally, missing data can bias studies of outage and health outcomes using New York State dataset and PowerOutages.us data. Both datasets contain substantial missingness - some counties in the poweroutages.us data are missing 70% of observations. In the poweroutages.us dataset, data are missing for entire utilities, because those utilities did not have a website. Data are also missing because utility websites were offline or unscrapable for long periods of time (months or years). In these cases, interpolating missing values is near impossible because there is little to no existing data from which to extrapolate.

To reduce bias, researchers could exclude counties that are missing more than a percentage of observations from epidemiological studies of outage exposure and health outcomes using these data. To do this, researchers must identify the threshold/percentage of observations missing at which missing data begins to severely bias effect estimates, and exclude counties missing more data than this threshold. In the following simulation, we aimed to find this threshold.

**Simulation design**

**Overview**

We designed a simulation representing an epidemiological study measuring the association between power outage exposure and hospitalization rates. This study was meant to mimic a study that could be conducted using poweroutages.us data. The outcome of ‘hospitalizations’ is intentionally vague, and could be any health outcome hypothesized to be caused by power outages. We simulated daily binary power outage exposure for 100 US counties for one year, and daily county-level hospitalization rates for these counties for the same time period. We generated effect estimates of power outage exposure on hospitalization in an unbiased scenario, and then in scenarios representing incorrect assumptions about health relevant duration, and scenarios including missing data.

**Exposure and outcome data**

We generated one year of county-hour power outage exposure data for 100 simulated counties. We populated each of the 100 counties with counts of electrical customers served, drawn from the empirical distribution of customers served by county in the poweroutages.us dataset. We then drew hourly county-level proportions of customers without power from the poweroutages.us dataset, and converted these into hourly counts of customers without power by county. Using proportions allowed us to assign reasonable numbers of customers without power to counties with varying numbers of customers served.

We chose a health-relevant duration of power outage, 8+ hours. This was arbitrary – in a real study, the health-relevant duration would depend on the actual outcome being studied, and how power outages were thought to cause that outcome. We applied our definition of power outage to these hourly counts of customers without power to identify county-days exposed to 8+ hour power outage. This produced a one-year daily timeseries of binary power outage exposure data for each county.

We then generated outcome data based on this exposure data. We drew hospitalization counts for each county-day based on the total number of customers served in a county from a Poisson distribution with a base rate of 0.1%. County-days that were exposed to 8+ hour outage received a 1% hospitalization rate increase (for a total hospitalization rate of 0.101%). This produced one-year time series of daily hospitalization rates for each county.

**Simulation study design**

First, we used a base case/unbiased scenario to estimate the true simulated effect of county-day 8+ hour power outage exposure on county-level hospitalization counts. We used a case-crossover design with a conditional quasi-Poisson model[[30]](#endnote-30). Within each county, we chose control days for each day with non-zero hospitalization count (each case day), and included case and control days in a quasi-Poisson model relating power outage exposure to hospitalization rates. In this model, we included an offset for customers served by county, and a fixed effect for county. We repeated this simulation, including exposure data creation, outcome data creation, modeling, 100 times.

We also repeated the simulation using a different study design, which we used to test if the simulation results were sensitive to study design. We implemented a simplified difference-in-differences design. We used the same exposure and outcome data generated for the case-crossover simulation. For each day exposed to power outage, we chose a control day not exposed to power outage from another county. We ran a Poisson model including all these case and control days from all 100 counties. We used exposure and outcome data created 100 times for the case-crossover design to repeat the difference-in-differences analysis 100 times.

**Testing wrong assumptions about health relevant duration:**

To model bias due to incorrect assumptions about the health-relevant duration of power outage, we created two additional power outage exposure datasets for each simulated county. Using the measurement strategy above, we identified county-day exposed to 4+ hour and 12+ hour duration power outages (customers without power counts > 10% of total customers for 4+ or 12+ consecutive hours). We generated two additional datasets of outcome data for each of the 100 counties, increasing the hospitalization rates when counties were exposed to either 4+ or 12+ power outages, rather than 8+ hour power outages. We used the same hospitalization rate of 0.1%, and a 1% rate increase on power outage exposed days.

We created a simulation meant to represent a researcher making wrong assumptions about the health-relevant duration of power outage. In this scenario, the true health relevant duration of power outage, causing increased hospitalizations, was 4+ or 12+, but the researcher had assumed 8+ power outages were relevant and identified these in the data. To represent this case, we paired exposure data indicating when counties were exposed to 8+ hour power outages with outcome data generated based on 4+ and 12+ hour exposure data, where days exposed to either 4+ or 12+ power outages had a 1% higher hospitalization rate.

We implemented the same study design as above to generate effect estimates for this scenario: a case-crossover design using conditional quasi-Poisson models. For each county and scenario (8+ hour exposure data paired with outcome data generated based on 4+ outages, and 8+ hour exposure data paired with outcome data based on 12+ hour power outages) we chose control days for each day with non-zero hospitalization count. We ran quasi-Poisson models to generate effect estimates for each of the two mismatched scenarios.

We repeated these two simulations 100 times each.

We also repeated the two simulations an additional 100 times using a difference-in-differences design, to test if results were sensitive to study design. Here, for each day exposed to power outage, we chose a control day not exposed to power outage from another county. We used those case and control days in a Poisson model to generate effect estimates for each of the two mismatched scenarios.

We calculated bias in the 4+ and 12+ hour cases using the absolute difference between the estimated effects and simulated effects (*𝛽*ˆ−*𝛽*; *𝛽*ˆ is the estimated effect and *𝛽* is the simulated effect) in each of the 100 models for each case using the case-crossover study design (Figure 1). We also calculated bias the same way for each of the 100 difference-in-differences models.

**Testing bias from missing data:**

To test bias from missing exposure data, we created four additional exposure datasets for each of the 100 simulated counties, each with an increasing percentage of missing observations (10%, 30%, 50%, 70%) in each dataset. To create missingness, we randomly removed county-hour observations from the original dataset until the correct percentage of observations were missing. We treated missing observations as though they indicated no power outage exposure (0 customers without power), since this is the average value of customers without power by hour in the poweroutages.us dataset, and it would be impossible to interpolate values more accurately in the real datasets. We applied our definition of power outage exposure to these four datasets with missingness to create daily binary power outage exposure data for 1 year for the 100 counties based on a power outage duration of 8+ hours.

We then modeled the relationship between 8+ hour power outage exposure measured in each of the four datasets with missing data (10% - 70% missing data) and all-cause hospitalization counts generated based on a 8+ hour power outage exposure in the complete dataset with no missingness in the 100 counties. We used the case-crossover study design as above and a difference-in-differences design as above. We repeated the simulations 100 times.

We calculated bias in each of the four cases with increasing missingness again using the absolute difference between the estimated effects and simulated effects (*𝛽*ˆ−*𝛽*; *𝛽*ˆ is the estimated effect and *𝛽* is the simulated effect) in each of the 100 models for each case using the case-crossover study design and difference-in-differences design (Figure 1).

**Results:**

**Discussion:**

1. <https://www.eia.gov/todayinenergy/detail.php?id=54639#:~:text=U.S.%20electricity%20customers%2averaged%20seven%20hours%20of%20power%20interruptions%20in%202021&text=On%20average%2C%20U.S.%20electricity%20customers,hour%20less%20than%20in%202020> [↑](#endnote-ref-1)
2. Casey JA, Fukurai M, Hernández D, Balsari S, Kiang MV. Power Outages and Community Health: a Nar- rative Review. Curr Envir Health Rpt. 2020;7(4):371-383. doi:10.1007/s40572-020-00295-0 [↑](#endnote-ref-2)
3. Wang, Shuai, and Ralf Toumi. "Recent migration of tropical cyclones toward coasts." *Science* 371.6528 (2021): 514-517. [↑](#endnote-ref-3)
4. National Oceanic and Atmospheric Administration. Atlantic Hurricane season takes infamous top spot for busiest on record. Published 2020. Accessed November 24, 2020. <https://www.noaa.gov/news/2020-atlantic-hurricane-season-takes-infamous-top-spot-for-busiest-on-record>

   [↑](#endnote-ref-4)
5. Li, Lin, and Pinaki Chakraborty. "Slower decay of landfalling hurricanes in a warming world." *Nature* 587.7833 (2020): 230-234. [↑](#endnote-ref-5)
6. Casey, J. A., Fukurai, M., Hernández, D., Balsari, S. & Kiang, M. V. Power Outages and Community Health: a Narrative Review. *Current Environmental Health Reports* **7**, 371-383, doi:10.1007/s40572-020-00295-0 (2020). [↑](#endnote-ref-6)
7. Chrobak, U. *The US has more power outages than any other developed country. Here’s why*, <<https://www.popsci.com/story/environment/why-us-lose-power-storms/>> (2020). [↑](#endnote-ref-7)
8. McBride, J. & Siripurapu, A. How Does the U.S. Power Grid Work? (2021). [↑](#endnote-ref-8)
9. Lindstrom A, Hoff S. U.S. electricity customers experienced eight hours of power interruptions in 2020. U.S. Energy Information Administration. Published November 10, 2021. Accessed January 11, 2022. <https://www.eia.gov/todayinenergy/detail.php?id=50316> [↑](#endnote-ref-9)
10. Casey, Joan A., et al. "Trends from 2008 to 2018 in electricity-dependent durable medical equipment rentals and sociodemographic disparities." *Epidemiology* 32.3 (2021): 327-335. [↑](#endnote-ref-10)
11. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-11)
12. Nunes, Ana Raquel. "General and specified vulnerability to extreme temperatures among older adults." *International journal of environmental health research* 30.5 (2020): 515-532. [↑](#endnote-ref-12)
13. Haman, François, et al. "Human vulnerability and variability in the cold: Establishing individual risks for cold weather injuries." *Temperature* 9.2 (2022): 158-195. [↑](#endnote-ref-13)
14. Andresen, Adam X., et al. "Understanding the social impacts of power outages in North America: a systematic review." *Environmental Research Letters* 18.5 (2023): 053004. [↑](#endnote-ref-14)
15. Andresen, Adam X., et al. "Understanding the social impacts of power outages in North America: a systematic review." *Environmental Research Letters* 18.5 (2023): 053004. [↑](#endnote-ref-15)
16. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-16)
17. Casey, Joan A., et al. "Power outages and community health: a narrative review." *Current environmental health reports* 7 (2020): 371-383. [↑](#endnote-ref-17)
18. Do, Vivian, et al. "Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA." *Nature communications* 14.1 (2023): 2470. [↑](#endnote-ref-18)
19. Eckstrom, Samuel, et al. "Outing power outages: real-time and predictive socio-demographic analytics for New York City." *2022 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2022. [↑](#endnote-ref-19)
20. Anderson GB, Bell ML. Lights Out: Impact of the August 2003 Power Outage on Mortality in New York, NY. Epidemiology. 2012;23(2):189-193. doi:10.1097/EDE.0b013e318245c61c [↑](#endnote-ref-20)
21. Xiao, Jianpeng, et al. "Increased risk of multiple pregnancy complications following large-scale power outages during Hurricane Sandy in New York State." *Science of the Total Environment* 770 (2021): 145359. [↑](#endnote-ref-21)
22. Do, Vivian, et al. "Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA." *Nature communications* 14.1 (2023): 2470. [↑](#endnote-ref-22)
23. Casey, Joan A., et al. "Power outages and community health: a narrative review." *Current environmental health reports* 7 (2020): 371-383. [↑](#endnote-ref-23)
24. Do, Vivian, et al. "Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA." *Nature communications* 14.1 (2023): 2470. [↑](#endnote-ref-24)
25. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-25)
26. Do, Vivian, et al. "Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA." *Nature communications* 14.1 (2023): 2470. [↑](#endnote-ref-26)
27. <https://poweroutage.us/about> [↑](#endnote-ref-27)
28. <https://www.eia.gov/electricity/> [↑](#endnote-ref-28)
29. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-29)
30. [↑](#endnote-ref-30)