Intro:

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. extreme heat events, wind, wildfires, hurricanes, and other severe weather[[3]](#endnote-3),[[4]](#endnote-4),[[5]](#endnote-5) --- the most common causes of power outages. 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 temperatures, precipitation, and wind now frequent with climate change. As a result, US electrical customers experienced an average of 8 hours without power in 2020, the longest duration on record[[8]](#endnote-8).

Power outages affect health in vulnerable populations such as older adults[[9]](#endnote-9), people who use life-sustaining electricity-dependent medical equipment[[10]](#endnote-10), and children[[11]](#endnote-11). Older adults may be susceptible to stroke, myocardial infarction, and other cardiorespiratory health effects from heat or cold exposure, when heating or air conditioning stops working during an outage[[12]](#endnote-12),[[13]](#endnote-13),[[14]](#endnote-14). These effects may be stronger when outages are caused by extreme heat or disasters. Loss of electricity can be life-threatening for people who use life-sustaining electricity-dependent medical devices such as at-home ventilators and oxygen tanks[[15]](#endnote-15). In children, outages increase injuries from accidents related to generator and natural gas use[[16]](#endnote-16). Outages also increase pediatric asthma emergencies, due to heat and humidity exposure without air conditioning[[17]](#endnote-17).

Despite the health risks of power outage, lack of data describing power outage exposure has limited research in this area[[18]](#endnote-18),[[19]](#endnote-19). Only one dataset describes outage exposure across space and time in New York state[[20]](#endnote-20), and most studies of outage exposure and health outcomes rely on this dataset. The remaining studies of power outage have relied on large-scale events such as hurricanes or disasters as a surrogate for power outage exposure[[21]](#endnote-21),[[22]](#endnote-22). These studies consider everyone in a city or county exposed to the 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 do not estimate exposure-response relationships between power outage exposure and health outcomes. Disentangling the effect of outage from disaster and quantifying the severity of health effects from power outage is critical to prevent hospitalizations and deaths from outages.

In our previous work, we created a new national dataset of 10-minute resolution power outage exposure in sub-county spatial units for the continental United States[[23]](#endnote-23). We used this dataset to describe power outage exposure by region and social vulnerability, finding that outages were more common in the southeast and northeast US, and with high outage incidence and high social vulnerability co-occurring most frequently in the southeastern US. This dataset will allow us to characterize exposure-response relationships between power outage and health outcomes nationally, by region, and within vulnerable populations.

However, major challenges with exposure assessment remain in using this newly available data. First, there is no standard or widely used strategy to measure power outage exposure in the literature[[24]](#endnote-24). A single strategy to describe power outage exposure would allow comparability and aggregation of results across studies. Second, any definition of outage exposure relative to an outcome will depend on the health-relevant length of a power outage: the length at which an outage begins to cause health effects. There is no literature on the clinically relevant length of power outage[[25]](#endnote-25),[[26]](#endnote-26) with respect to any outcome. Incorrect assumptions about the clinically relevant length of power outage could substantially bias the results of an epidemiological study. Finally, both the new national dataset and existing New York State data are missing large percentages of observations[[27]](#endnote-27),[[28]](#endnote-28). This missingness could also substantially bias results of an epidemiological study of power outage and any health outcome.

In this paper we will address these issues with power outage exposure assessment by developing a strategy for measuring power outage exposure. Then, we will run simulations to test how assumptions about clinically relevant length 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 effects.

Methods:

Power outage exposure measurement strategy:

Existing data:

In our previous work, we created a national dataset of power outage exposure. We purchased raw power outage data covering the continental US for the years 2018-2020 from poweroutages.us. Poweroutages.us scrapes data from utility company website APIs, designed to be used by utility customers to check for power outages in their area. Poweroutages.us scraped counts of customers without power from these websites in real time, every ten minutes, from 2018-2020[[29]](#endnote-29). The resulting data contains 10-minute resolution counts of customers without power all served by the same utility in a sub-county unit. Utilities define a ‘customer’ as a grid connection, which can correspond to a household, apartment building, or business[[30]](#endnote-30). A sub-county unit can be an entire county, city (where there are possibly multiple cities in a county), or neighbourhood (where there are possibly multiple neighbourhoods in a city, in turn nested in a county). Many utilities can serve the same location, so these sub-county units were not necessarily geographically distinct. Two houses next to each other might be in two different spatial units in the power outage data if they were served by different utilities.

The sub-county unit level data is a time series of ten-minute intervals covering 3 years, where there is an estimate of the number of customers without power in a sub-county unit in every 10-minute interval. The counts of customers out do not necessarily track the same customers: if 10 people are reported without power in two subsequent 10-minute periods in a subcounty unit, 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 interval. New York State power outage data are structured similarly.

Because all subcounty units were nested inside counties, we were able to aggregate 10-minute counts of customers without power to the county level.

Other studies have defined binary power outage exposure over a spatial unit (whether there was an outage, Y/N, in a county or zip code etc.) by considering an outage “on” if the percentage of customers without power in a spatial unit exceeds a threshold[[31]](#endnote-31)[[32]](#endnote-32)[[33]](#endnote-33)[[34]](#endnote-34). There is exposure misclassification inherent in this definition: when the spatial unit is ‘exposed’, some households in the spatial unit will be without power and others will not. Studies have dealt with this exposure misclassification by conducting sensitivity analyses varying the threshold 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[[35]](#endnote-35). As the threshold increases, the specificity of this definition of power outage increases.

For this simulation study and future epidemiological studies using the poweroutages.us dataset, we developed the following binary daily county-level definition of power outage. This definition is similar to previous definitions in the literature. We considered a county-hour exposed to power outage if the percentage of customers without power in county *i* during hour *j* exceeded 0.5% of the customers served in county *i*.[[36]](#footnote-1) We then aggregated this hourly exposure to the daily level: we considered a county-day as exposed if there were at least 8 consecutive hours of ‘power outage on’ (customers without power percentages > 0.5% 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 8 hours ended on that county-day.

By using 8+ hours as the threshold in this definition, we assumed that the clinically relevant length of power outage at the county level was 8 hours. With this definition, we can run a simulation testing how much bias may be introduced into an epidemiological study if this assumption is incorrect. In real studies, we could vary the percentage of customers without power in sensitivity analyses as other studies have done.

Simulation design:

We designed a simulation representing an epidemiological study measuring the association between power outage exposure and all-cause hospitalization rates by county-day, though this outcome is standing in for many potential outcomes of interest related to power outages. In this hypothetical study, daily binary power outage exposure is measured in 100 US counties for 1 year. Daily county-level hospitalization rates are also measured over 1 year, and the study aims to estimate the effect of county-level binary power outage exposure on county-level hospitalization rates using a difference-in differences study design. We used this setup to test how much incorrect assumptions about the length of clinically relevant power outage, and missing data, would bias the results of an epidemiological study of power outage and a health outcome.

Real power outage data:

We wanted to simulate power outage data resembling the real power outages we have assembled and plan to use in future studies. To assemble this dataset, in our previous work, we purchased raw power outage data covering the continental US for the years 2018-2020 from poweroutages.us. These data come from public websites maintained by utility companies, designed to be used by utility customers to check for power outages in their area. Poweroutages.us scraped counts of customers without power from these websites in real time, every ten minutes, from 2018-2020[[37]](#endnote-36). The resulting data contains 10-minute resolution counts of customers without power all served by the same utility in a sub-county unit. Utilities define a ‘customer’ as a grid connection, which can correspond to a household, apartment building, or business[[38]](#endnote-37). A sub-county unit can be an entire county, city (where there are possibly multiple cities in a county), or neighbourhood (where there are possibly multiple neighbourhoods in a city, in turn nested in a county). Many utilities can serve the same location, so these sub-county units were not necessarily geographically distinct. Two houses next to each other might be in two different spatial units in the power outage data if they were served by different utilities.

The sub-county unit level data is a time series of ten-minute intervals covering 3 years, where there is an estimate of the number of customers without power in a sub-county unit in every 10-minute interval. The counts of customers out do not necessarily track the same customers: if 10 people are reported without power in two subsequent 10-minute periods in a subcounty unit, 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 interval.

Because all subcounty units were nested inside counties, we were able to aggregate 10-minute counts of customers without power to the county level.

Measuring power outage exposure:

Other studies have defined binary power outage exposure over a spatial unit (whether there was an outage, Y/N, in a county or zip code etc.) by considering an outage “on” if the percentage of customers without power in a spatial unit exceeds a threshold[[39]](#endnote-38)[[40]](#endnote-39)[[41]](#endnote-40)[[42]](#endnote-41). There is exposure misclassification inherent in this definition: when the spatial unit is ‘exposed’, some households in the spatial unit will be without power and others will not. Studies have dealt with this exposure misclassification by conducting sensitivity analyses varying the threshold 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[[43]](#endnote-42). As the threshold increases, the specificity of this definition of power outage increases.

The length of power outages matter. Studies have defined daily binary power outage exposure over a spatial unit by considering a unit exposed if there were 8+ consecutive hours where the percentage of customers out exceeded a threshold such as 10% within a 24-hour period[[44]](#endnote-43)[[45]](#endnote-44). However, there is no literature on the clinically relevant length of power outage. 8+ hours has been used because batteries for most electricity-dependent medical equipment last 8 hours[[46]](#endnote-45). During a power outage, electricity-dependent medical device users might experience adverse health effects without their equipment immediately after losing power to the equipment[[47]](#endnote-46). Without air conditioning or heat, indoor temperatures may also begin to change over the course of 8 or more hours[[48]](#endnote-47),[[49]](#endnote-48).

The clinically relevant length of outage depends on the health outcome being studied. In reality, for medical-device related emergencies, pediatric asthma and injuries, and cardiorespiratory health effects in older adults, some individuals may experience health effects as soon as a power outage begins. As an outage lasts longer, the likelihood of adverse health effects increases, and we hypothesize that there are threshold effects for some outcomes (ex: 8+ hours for medical equipment related emergencies, or some number of hours for heat).

There is a difference between the clinically relevant length of power outage for an individual, and the clinically relevant ‘length of outage’ in a spatial unit. If the number of customers out in a spatial unit exceeds 10% for 8 hours, this does not mean that 10% of customers are without power for 8 hours consecutively. Some customers could have power restored after 4 hours, and others could experience an outage 4 hours into the 8-hour window, and contribute to the percentage of customers without power. If the number of customers out in a spatial unit exceeds 10% for 8 hours, this may indicate that many individuals in the spatial unit are without power for approximately 8 hours, which may in turn produce health effects. Again, in reality, as outages (as measured at the spatial unit level) last longer, the likelihood of individuals in that spatial unit experiencing adverse health effects increases. We hypothesize that there are threshold effects at certain outage lengths for certain outcomes at the spatial unit level as well as the individual level for the same reasons.

For this simulation study and future epidemiological studies using the poweroutages.us dataset, we developed the following binary daily county-level definition of power outage. This definition is similar to previous definitions in the literature. We considered a county-hour exposed to power outage if the percentage of customers without power in county *i* during hour *j* exceeded 0.5% of the customers served in county *i*.[[50]](#footnote-2) We then aggregated this hourly exposure to the daily level: we considered a county-day as exposed if there were at least 8 consecutive hours of ‘power outage on’ (customers without power percentages > 0.5% 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 8 hours ended on that county-day.

By using 8+ hours as the threshold in this definition, we assumed that the clinically relevant length of power outage at the county level was 8 hours. With this definition, we can run a simulation testing how much bias may be introduced into an epidemiological study if this assumption is incorrect. In real studies, we could vary the percentage of customers without power in sensitivity analyses as other studies have done.

Simulation on the clinically relevant length of power outage:

In this portion of the simulation, we aimed to test how much incorrect assumptions about the length of clinically relevant power outage would bias the results of an epidemiological study. At the spatial unit level, if ‘power outage exposure’ as defined above (at least X% of customers without power for Y number of hours) produces the most health effects when Y >= 8, how much would misidentifying the length of outage bias the results of an epidemiological study? If it was instead 4+ hour outages or 12+ hour outages (as measured at the spatial unit level) that were clinically relevant and mattered for health, and we cleaned the data as if it were 8+ hour outages that mattered, we would be introducing non-differential exposure misclassification.

To test how much this misclassification could bias results, we modeled two scenarios where the ground truth was that 4+ or 12+ power outages caused health effects, but the researchers conducting the study incorrectly assumed that 8+ outages caused health effects. In this simulation, we created outcome data – hospitalization data – where we increased the hospitalization rate when a county-day was exposed to a 4+ hour outage or a 12+ hour outage. Then, we modelled the relationship between the outcome data generated based on the 4+ hour exposures and 12+ hour exposures, and the exposure data of 8+ hour power outages. Finally, we evaluated by how much this exposure misclassification biased the effect estimates from these simulated studies.

Data preparation for simulation on clinically relevant length of power outage:

To create simulated data for an epidemiological study measuring the association between binary daily power outage exposure and daily county-level all-cause hospitalization rates, we started by generating exposure data for 100 simulated counties. We created power outage exposure data as similar as possible to the real poweroutages.us data. We assigned each county a number of sub-county areas. We determined the number of sub-county areas by drawing from the empirical distribution of sub-county areas in the poweroutages.us dataset. We populated each sub-county area with a total number of simulated customers served, again drawn from the empirical distribution of customers in sub-county areas in the real poweroutages.us dataset. We then generated ten-minute counts of customers without power for each sub-county unit for one year. The counts of customers without power at each 10-minute interval were also drawn from the empirical distribution of the proportion of customers without power in the poweroutages.us data, then multiplied by the customers served in the sub-county unit. We constructed our simulated data from the proportion of customers without power because of very different counts of customers served by sub-county unit nationwide.

We aggregated these sub-county 10-minute counts of customers without power to the hourly level. We also aggregated hourly counts of customers out to the county level, by summing all the customers served in each sub-county area and the customers out in each hour in each sub-county area (Figure 1).

We used the definition of power outage described above to mark county-hours exposed or unexposed to power outage. We marked a county-hour exposed if the percentage of customers without power in county *i* during hour *j* exceeded 0.5% of the customers served in county *i*. Finally, we aggregated to the daily level: we considered a county-day as exposed if there were 8 consecutive hours of ‘power outage on’ (customers without power percentages > 0.5% 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 8 hours ended on that county-day. This process produced a one-year time series indicating if there was an 8+ hour power outage on each day for each of the 100 simulated counties.

Modelling effects of power outage on hospitalization and measuring bias:

First, to establish a ground truth case representing an unbiased scenario where exposure was measured correctly, we generated simulated outcome data of all-cause hospitalization counts by county-day based on the simulated 8+ hour power outage exposure data. For each of the 100 simulated counties, we drew hospitalization counts for each county-day based on the total number of customers living 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% rate increase (for a total hospitalization rate of 0.101%). This produced one-year time series of daily hospitalization rates for each of the 100 counties.

To generate a ground-truth estimate for the effect of an 8+ hour power outage on daily county-level hospitalization counts, for each day exposed in a county, we chose control days that were more than 1 week away from the exposed day, but within 4 weeks of the exposed day, matching on the day of week. This process was meant to approximate a study using an augmented difference-in-differences design, which could be used to compare hospitalization rates in counties exposed to power outages with those not exposed. We used these exposed and control dates in Poisson models. We used one Poisson model per county (for a total of 100 models) to model the relationship between the binary 8-hour power outage exposure and daily hospitalization counts, with an offset for the number of customers in a county. These models represented an unbiased scenario where exposure was measured correctly – 8+ hr exposure caused an increase in hospitalizations. This was reflected correctly in the exposure data used in the simulation.

To model exposure misclassification due to incorrect assumptions about the clinically relevant length of power outage, we first created two additional exposure datasets for each of the 100 counties, marking a county-day as exposed if there was either a 4+ hour outage, or 12+ hour outage (customers without power counts > 0.5% of total customers for 4+ or 12+ consecutive hours), instead of an 8+ hour outages.. We generated two additional datasets of outcome data for each of the 100 counties (simulated all-cause hospitalization data) based on the same hospitalization rate of 0.1%, and a 1% rate increase on days with 4+ hour and 12+ hour power outages.

We induced exposure misclassification by pairing 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 used two Poisson models per county (for a total of 100 models for each scenario) to model the relationship between the binary 8+ hour power outage exposure and daily hospitalization counts based on the 4+ hour exposure data, and 12+ hour exposure data. Again, we used an offset for the number of customers in a county. In the case where 4+ hours was the relevant exposure, we would have classified many exposed county-days as unexposed, whereas if 12+ hours was the relevant exposure, we would have classified many unexposed county-days as exposed when we used an 8+ hour duration for analysis.

We calculated bias due to exposure misclassification in the 4+ and 12+ hour cases using the absolute difference between the estimated effects and simulated effects in each of the 100 models for each case (*𝑙𝑛*(*𝛽*ˆ)−*𝑙𝑛*(*𝛽*); *𝛽*ˆ is the estimated effect and *𝛽* is the simulated effect). We also calculated coverage – the proportion of simulations where the 95% confidence intervals included the simulated effect (Figure 2).

Simulation on bias due to missing data:

Many counties in the poweroutages.us dataset are missing substantial amounts of data. Some are missing up to 70% of observations. This missingness has the potential to bias effect estimates from an epidemiological study of power outage exposure and any health outcome. Counties with no or little missing data contribute information towards the overall effect estimate, while including counties with substantial amounts of missing data in an analysis could introduce bias.

Substantial missingness in the poweroutages.us dataset happens in two ways. In some cases, data are missing for an entire sub-county unit: the unit is never included in the dataset. This may happen because a sub-county unit is located in a rural area, and is served by a small co-operative utility without a website. Because poweroutages.us data is scraped from websites, no data – no average value of customers without power, or even the number of customers served by the utility in the subcounty unit – is included in the dataset. Because there is no information at all included in the dataset about these missing areas, interpolating the number of customers without power in missing subcounty units is near impossible to do with any accuracy.

Data may also be missing from the poweroutages.us website because utility websites may be offline or inaccessible for long periods of time (months or years). People living in the areas served by utilities without active websites are not represented in the dataset during the time periods when websites are not accessible. In this case, interpolating missing values is also near impossible, since values in a months or years-long period of missingness can’t be interpolated from prior observations with any accuracy.

In this simulation, we treat missing data as if it indicates that no exposure: that there are no customers without power at any time in the missing subcounty units and missing time periods. Zero is by far the most common value, and the average value, of customers without power in the non-missing subcounty level poweroutages.us data. If data are missing at random, this strategy for addressing missing data means any bias from missing data would be towards the null.

To reduce bias due to missing data in an epidemiological study of power outage exposure and a health outcome using the poweroutages.us dataset, researchers could exclude counties that are missing more than a threshold percentage of observations. Within counties with large amounts of missing data, treating missing data as if it indicates no exposure, effect estimates are likely to be very biased towards the null. Counties with no or little missing data contribute information towards the overall effect estimate, and within these counties, bias is likely small. Excluding counties with high missingness could result in less biased overall effect estimates. To do this, researchers must identify the threshold at which missing data in a county begins to severely bias effect estimates, and use this threshold to determine which counties to exclude from an analysis.

In this simulation, we aimed to find this threshold. We simulated an epidemiological study of daily county-level power outage exposure and daily county-level hospitalizations, as in the previous simulation. We aimed to estimate how increasing amounts of missing exposure data would bias the effect estimates of the study. We created exposure datasets missing 10%, 30%, 50%, and 70% of observations, and outcome data based on a complete dataset. We conducted an analysis pairing each of the incomplete datasets with outcome data based on the complete dataset, and assessed bias in each of the four cases.

Data preparation for simulation on missing data:

We generated power outage exposure data for 100 simulated counties as in the previous simulation. We aggregated the simulated data from sub-county units to the county level, and identified days exposed and unexposed to 8+ power outages. As in the previous simulation, this process produced a one-year time series indicating if there was an 8+ hour power outage on each day for each of the 100 simulated counties.

As in the previous simulation, we generated simulated outcome data for each of these 100 counties. We generated simulated all-cause hospitalization counts by day and county based on the simulated 8+ hour exposure data. We drew hospitalization counts for each county-day based on the total number of customers living in a county from a Poisson distribution with a base rate of 0.1%. County-days that met the 8-hour outage definition received a 1% rate increase (for a total hospitalization rate of 0.101%). This produced one-year time series of daily hospitalization rates for each of the 100 counties.

As in the previous simulation, to generate a ground-truth estimate for the effect of an 8+ hour power outage on daily county-level hospitalization counts, for each day exposed to a power outage in a county, we chose control days that were more than 1 week away from the exposed day, but within 4 weeks of the exposed day, matching on the day of week. As above, this process was meant to approximate a study using an augmented difference-in-differences design comparing hospitalization rates in counties exposed to power outages with those not exposed. Again, we used these exposed and control dates in Poisson models. We used one Poisson model per county (for a total of 100 models) to model the relationship between the binary 8+ hour power outage exposure and daily hospitalization counts, with an offset for the number of customers in a county. In this simulation, these models represented an unbiased scenario where no data was missing.

To model missing exposure data, we created four additional exposure datasets for each of the 100 counties, each with an increasing percentage of missing observations (10%, 30%, 50%, 70%). To create missingness, before aggregating exposure datasets from the 10-minute subcounty level, we randomly removed 10-minute sub-county level observations from the original dataset until the correct percentage of observations were gone. We then aggregated these four datasets to the county-day level as above, treating missing observations as if they indicated no power outage exposure.

We then modeled the relationship between exposure in each of the four datasets with missing data (10% - 70% missing data) and all-cause hospitalization counts generated based on a complete dataset in each of the 100 counties. We chose exposed and control days as we did for the ground truth scenario. We used four Poisson models per county (one each for 10%, 30%, 50%, and 70% missingness respectively for a total of 100 models per dataset) to model the relationship between the binary 8+ hour power outage exposure based on the datasets with increasing missingness, and daily hospitalization counts based on the complete 8+ hour exposure data, again with an offset for the number of customers in a county.

We calculated bias due to missing data for each of the four cases using the absolute difference between the estimated effects and simulated effects in each of the 100 models (*𝑙𝑛*(*𝛽*ˆ)−*𝑙𝑛*(*𝛽*); *𝛽*ˆ is the estimated effect and *𝛽* is the simulated effect). As in the previous simulation, we also calculated coverage – the proportion of simulations where the 95% confidence intervals included the simulated effect (Figure 2).

Possibly for the supplement:

To remove the correct amount of data, we quantified the percentage of missing data in the following way: If no power outage exposure data was missing from a county, then the data should have information on all the customers served for all of the hours in a year. The total amount of data missing from a county can be represented as a percentage of that total customer-hours in county *i*. Customer-hours present in county *i* should be:

*[customers served in county i] \* [number of hours in the year] = N customer-hours*

When we removed data, we removed observations before aggregating the data. We removed counts of customers out in a sub-county unit for one 10-minute time period. Once enough observations were missing, we aggregated the dataset as above to get the county-day timeseries of power outage exposure used in the simulation.

Needs to go somewhere but I’m not sure where:

Caveat: it’s likely that not all the data missing from poweroutages.us are missing at random. Some may be – some utilities may not have websites simply because they haven’t developed them, or utility websites may be undergoing maintenance or development, which could lead to random missingness. However, a lot of missingness is likely patterned: for example, small, rural utilities are much less likely to have a website or API, meaning that many of those utilities may not be in the poweroutages.us dataset. It’s also possible that utility websites and APIs crash during extremely large outages, meaning that there may be missing data when counts of customers out are extremely high. Additionally, in both these cases, the missingness may be related to health outcomes influenced by power outage. For the outcome of hospitalization, rural counties may have different hospitalization rates than urban ones. Large power outages are often caused by natural disasters, which may also affect hospitalization rates. However, to evaluate the potential effects of missingness not at random, we would have to speculate about so many different simulation parameters with almost no information to inform our assumptions that such a simulation would not be informative. We therefore limit ourselves to an investigation of missingness that we assume is random.

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. 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-8)
9. 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-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. 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-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. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-17)
18. Casey, Joan A., et al. "Power outages and community health: a narrative review." *Current environmental health reports* 7 (2020): 371-383. [↑](#endnote-ref-18)
19. 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-19)
20. 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-20)
21. 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-21)
22. 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-22)
23. 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-23)
24. Casey, Joan A., et al. "Power outages and community health: a narrative review." *Current environmental health reports* 7 (2020): 371-383. [↑](#endnote-ref-24)
25. Casey, Joan A., et al. "Power outages and community health: a narrative review." *Current environmental health reports* 7 (2020): 371-383. [↑](#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. 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-27)
28. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-28)
29. <https://poweroutage.us/about> [↑](#endnote-ref-29)
30. <https://www.eia.gov/electricity/> [↑](#endnote-ref-30)
31. Dominianni, Christine, et al. "Health impacts of citywide and localized power outages in New York City." *Environmental Health Perspectives* 126.6 (2018): 067003. [↑](#endnote-ref-31)
32. Lin, Shao, et al. "The immediate effects of winter storms and power outages on multiple health outcomes and the time windows of vulnerability." *Environmental Research* 196 (2021): 110924. [↑](#endnote-ref-32)
33. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-33)
34. 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-34)
35. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-35)
36. This is smaller than the percentage thresholds described above, because counties are larger than the spatial units used in Northrop et al., meaning that the percentage of customers that need to be out for it to be a ‘substantial’ power outage is a smaller proportion of the population. [↑](#footnote-ref-1)
37. <https://poweroutage.us/about> [↑](#endnote-ref-36)
38. <https://www.eia.gov/electricity/> [↑](#endnote-ref-37)
39. Dominianni, Christine, et al. "Health impacts of citywide and localized power outages in New York City." *Environmental Health Perspectives* 126.6 (2018): 067003. [↑](#endnote-ref-38)
40. Lin, Shao, et al. "The immediate effects of winter storms and power outages on multiple health outcomes and the time windows of vulnerability." *Environmental Research* 196 (2021): 110924. [↑](#endnote-ref-39)
41. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-40)
42. 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-41)
43. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-42)
44. 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-43)
45. Northrop, Alexander J., et al. "Power outages and pediatric unintentional injury hospitalizations in New York State." *Environmental Epidemiology* 8.1 (2024): e287. [↑](#endnote-ref-44)
46. Casey, Joan A., et al. "Power outages and community health: a narrative review." *Current environmental health reports* 7 (2020): 371-383. [↑](#endnote-ref-45)
47. Mango, Marriele, Joan A. Casey, and Diana Hernandez. "Resilient Power: A home-based electricity generation and storage solution for the medically vulnerable during climate-induced power outages." *Futures* 128 (2021): 102707.

    [↑](#endnote-ref-46)
48. Sheng, Maggie, et al. "Assessing thermal resilience of an assisted living facility during heat waves and cold snaps with power outages." *Building and Environment* 230 (2023): 110001. [↑](#endnote-ref-47)
49. Attia, Shady, et al. "Resilient cooling of buildings to protect against heat waves and power outages: Key concepts and definition." *Energy and Buildings* 239 (2021): 110869. [↑](#endnote-ref-48)
50. This is smaller than the percentage thresholds described above, because counties are larger than the spatial units used in Northrop et al., meaning that the percentage of customers that need to be out for it to be a ‘substantial’ power outage is a smaller proportion of the population. [↑](#footnote-ref-2)