Memo to Joan and Marianthi

Power outage exposure and national Medicare hospitalizations

Project progress so far (Nov 6th, 2024):

Analysis plan recap/how it went:

* Unit of analysis: county-day
* Coverage: 69% of Medicare benes after excluding counties w bad power outage coverage, and counties w less than 500 benes.
* Time period: 2018
* Exposure: 8+ hour power outage, and also integer number of hours without power
  + According to simulation study removing 50% of data is a good strategy (can discuss this if needed). In short, I’d say that 25% of counties are missing 50% of data or more, and the rest are almost complete (mean completion 90%). When we exclude counties missing 50% of data or more, 25% of counties are missing 30% of data and the rest are more than 90% complete. Our simulation shows bias is minimal in the latter case. Proof is also in the pudding in that we have non-null results when we use this strategy (although this is not a good way to decide on this in general).
  + Customers served (denominators for customers within county used to calculate power outage) are from EIA downscaled to county using census households and establishments, not used as denoms in main model, just used to define outage.
  + I looked at different cut points for deciding when a county-hour has a power outage. (0.005 (initial one used in Vivian and dominianni/zhang), 0.01, 0.03, and 0.05), thresholds above this have very few power outages. So I went with these. We could consider doing analyses with percentile-based metrics or with 10% out, but I suspect these are good choices (QUESTION: do you guys agree?)
  + Decided 8+ hour outage affected 1% of county (0.01\*customers\_served) should be main analysis
  + Did sensitivity analyses on 12 durations and cut points – 4, 8, 12+ hrs, and 0.005-0.05
  + Checked autocorrelation of outages of all durations and cut points – it’s always about 0.2. same w continuous hours out so we don’t need a dlnm in either case
* Outcome: Medicare hosp in older adults (65+)
  + Have cleaned Medicare data and tabulated hospitalization counts by county-day
  + Calculating hospitalizations by filtering to older adults
  + Using beneficiaries as denominator for the number of people eligible for hospitalization in a county
  + Excluding counties w less than 500 benes. There is not really a threshold after which rates become ‘stable’, (I plotted this and checked), but this seems to be working ok? Not sure how to pick a threshold.
  + QUESTION: internet says there should be 46 million benes >= 65 in 2018, and the dataset I got includes 33 million 65 >=. Why, who to ask about this? NSAPH data cleaning mistake?
  + Using urgent and emergency hospitalizations as per our group conversation with Robbie and Dan about why we should use those, and bc this is in line w our theory
  + Aiming to capture all resp and cvd, using first five ICD codes. We initially excluded hem stroke and hypertension. I did a sensitivity analysis with these included, we should discuss. (QUESTION: discuss)
* Study design: conditional Poisson
  + Checked strength of weekday and seasonal effects, both are weak but present, and 2 month strata seem to be fine. Stratified by 2-month period and by day of week.
  + Controlling for time-varying confounders: temp, wind speed, precipitation with gridmet. Has county day total precip (yes it includes snow, no I have not been able to find the units despite searching for TOO LONG -\_-), average wind speed in m/s, and max temp and min temp.
  + Open question: what is the appropriate way to control for wind speed and precip? Have done some investigations, QUESTION: we need to talk about these.
  + Checked dispersion: 0.91 or similar on all models

Analyses so far:

* Main analyses: 8+ hour outages affecting 1% of pop and cvd and respiratory hosp, no hem no hyp – completed, results are preliminary before decision about controlling for confounders. (see plots) note I have done these controlling for wind speed and precip linearly and with a natural spline w 3dfs, and the results were the same.
* Sensitivity on all cut points and durations cvd and respiratory hosp, no hem no hyp – completed (see plots)
* Analyses with only resp and only cvd, as well as analysis with all hosp included hem and hyp – complete, as well as these on all cut points and durations (see plots)
* Also completed analyses on 8+ hour outages affecting 1% of pop and cvd, resp, and cvd + resp stratified by age and sex (see plots)

Analyses still to complete and progress on them:

* Effect modification analyses, by DME use, poverty, and hot and cold days
* DME use:
  + Got empower counts, calculated proportion DME users, found quartiles of DME users. Want to compare effects of power outage on hosp in 1st quartile vs 4th quartile
* Poverty:
  + got ACS 2015-2019 info on number of hh in counties and what proportion of the poverty income they make, in bins (0 – 0.5 of pov income, 0.5-1, etc.)
  + summed these counts to get what proportion of county makes < 1 times poverty income
  + found quartiles, compare 1st to 4th quartile.
* Hot and cold days:
  + Identified anomalously hot and cold days using same strategy alex Northrop et al used. Identified 15th and 85th percentile and also put absolute thresholds of 0 and 27 deg. If a day was outside these percentile thresholds and also outside the absolute thresholds, it was considered hot or cold
  + Calculated using prism data 1981-2010
  + I’m thinking we want to see if effects are stronger on these days, maybe separately for hot and cold days. Not sure how to do this though bc I think we may have a power issue (not enough hot and cold days in 2018). Have not checked yet about how many days there actually are. QUESTION: how should we do this?
* CONTINUOUS MEASURES
  + Question: how and what do we want to analyze with continuous measures??

Questions:

* Need to look at results together
* How do we want to control for confounding/handle our issues around not knowing how flexibly we should control for wind speed and precip
* How to calculate cumulative effect uncertainty, and how do we want to do this (which results do we plan to present?)
* Main question: what figures do we want in the paper, so that I can start making those with the results, and do we want age and sex stratified by resp and cvd separately, or do we want all hospitalizations lumped into a single outcome? Do we want to present cumulative effects or lags like we have now? do we want a figure showing our coverage of medicare benes, and do we want to highlight any urban areas? Table 1?