Wildfire disaster exposure and preterm birth analysis plan

Authors:

* Heather – first author, doing most of the things
* Joan
* Tim
* Marianthi
* Dana – last author?
* Ray? (could he be? That would make me happy)
* Anyone else from tim’s group?
* Need a student code reviewer. Vivian or Lauren?

Supervisors are providing advice and feedback by coming to meetings, and reading and revising the manuscript. Is Dana last author?

Study question:

How does wildfire disaster exposure influence risk of preterm birth?

Exposure data:

Lab wildfire dataset covering outcome data period

Outcome data:

Tim data on preterm births at the county level 2000-2019

Exposure assessment plan:

* Aim is to capture stress, evacuation, smoke, community disruption, power outage, healthcare disruption, housing issues related to wildfire disaster exposure.
* Need exposure at county level
* Going to use different buffers and popexposure to determine how much of each county population is exposed
* Buffer distances 5km, 10km, 20km from fire boundary
* Primary analysis 10km buffer
* If >50% of pop is exposed, 20% of pop is exposed, consider a county exposed
* Do sensitivity analyses on buffer distance
* Option to change buffer distances to get enough people exposed to do the study; tbh in EDA
* Exposure is going to be monthly. County-level monthly isn’t going to be great resolution, but it we do not have dates when the wildfire ends, and exposure to huge disasters likely doesn’t really ever end. So we’re going to have monthly exposure. Weekly exposure seems to fine a resolution. Also we want to have good outcome counts that are non-zero so doing things monthly will help with that
* Once exposed, exposed for two years
* Need to vary the washout period in a sensitivity analysis

Optional:

* Consider co-exposure to wildfire smoke and power outage along with disaster exposure
* Could look at places co-exposed to both a threshold of wf pm or county-level power outage co-occurring with the wildfire disaster. But that might be outside the scope of this paper
* Could also analyze only the top 100 most destructive fires or something

Outcome assessment:

* Pretty simple
* PTB rate
* Could also consider early preterm
* County-level monthly rate should be fine
* Need to consider any data issues with the birth dates/conception cohort issues where earlier shocks can influence PTB rate later

Optional:

* Disparities/SVI secondary analysis

Confounders:

* Heat, and maybe severe weather. We can use PRISM to estimate this but should talk to make sure we have data for Alaska and Hawaii since they are so exposed
* Day of week effects, population changes, changes in evacuation and warning procedures, changes in fire size and severity, and other things could affect the effect of wf on hospitalizations but none of these things do so differentially across the treatment timing, so they shouldn’t bias estimates

Analysis:

* DID analysis, staggered multi-region setup
* Estimate the group-time ATT with R package did
* Aggregate group ATTs to look for dynamic effects 1, 2, and 3 months after wildfire disaster exposure
* CIs are corrected for multiple comparisons + autocorrelation within region
* Three-month follow-up period (run sensitivity analyses on that, make it one vs three months as well)
* Two-year washout period (sensitivity on that as well, make it one and three years)

Analytic concerns and how we will address them:

Assumptions of staggered, multi-region DID:

* On average, trends in untreated or not-yet-treated regions are parallel to treated regions after controlling for confounders
* No anticipation
* No coincident changes
* No interference
* Correct model spec

Parallel trends:

* DID package has a thing built in where we will be able to test this assumption in the pre-treatment periods which is the best we can do
* I think that average parallel trends after controlling for temperature is plausible
* Will test using the DID built in pre-testing

No anticipation:

* I think this is a fair assumption and I can’t think of a way to test it.
* DID has a way of relaxing this if it’s super important to us but idk I think it’s fine. I don’t think people are anticipating the majority of our exposure

No coincident changes:

* For this to cause problems in the staggered multi-region DID analysis, you need something that is causing systematic changes in your outcome in the same direction in every treated region
* These are ‘confounders’
* Like temperature
* Which we are controlling for
* And I can’t think of another

Correct model spec:

* The model is pretty flexible and doubly-robust
* Need to learn more about what’s going on with this

No interference:

* Here I think we can do a sensitivity analysis where we exclude neighboring counties that were next to an affected county and see how this changes results

Other considerations:

* Might have issues with effect heterogeneity between regions or the affected regions being systematically different from those unaffected (this would be an exchangeability issue)
* Need to be careful to acknowledge that it’s the ATT
* Also might want to analyze things separately for different eras (2000-2005, 2006-2010, etc). if we’re interested in heterogeneity as fires got larger and smoke etc. is a bigger issue later in the study period
* What if we randomly assigned exposure as a negative control? That might be interesting
* Need to chop up the data into cross-sections with pre-treatment time, at least three months post treatment, and then a washout period. Should think about how to do this.

TO-DOs:

* Think about PTB rate stuff and birthday stuff and make sure there are no issues
* Make a dag?
* Decide what optional and sensitivity analyses to do. There are like 15 possible analyses laid out here. We do not want all of them!
* Learn about model
* What about regions next to each other having similar outcomes? Does that matter?