PEGS Study Plan

Insang Song Kyle P. Messier

Background/study rationale

Extreme weather events (EWE) have increasingly gained attention from research communities across environmental epidemiology, health geography, and public health in relation to climate change. Most studies assess the mental health effect of extreme weather by accounting for the extreme weather event as a single separate event. This practice may pose the possibility of either inflated or deflated effects on stress level as residential relocation and the length of residence could affect the heterogeneity in vulnerability and response to stress among individuals depending on adaptation or habituation. This study aims to account for residential history as an effect modifier of the main effect of exposure to the local intensity of extreme weather events on stress levels. Using the Personalized Environment and Gene Study data, we will examine the hypothesis by leveraging residential relocations derived from addresses at study dates in Personalized Environment and Gene Study (PEGS).

Study hypothesis(es) to be tested

We pose the following hypotheses:

- The stress level of individuals is associated with recent exposure to extreme weather events.
- The association between the stress level of individuals and the extreme weather events is mediated by cumulative exposure to EWE derived from residential history.

Specific aims and plans

We aim to estimate the mediation effect of the duration of residence to the effect of exposures to extreme weather events on the stress level of individuals. The main exposure of extreme weather events is our primary interest, and the duration of residence is our mediator. The mediation effect is examined after accounting for all potential confounders including demographic, socioeconomic, environmental, and genetic factors. The study consists of four steps. First, we will identify the extreme weather events that individuals have experienced in

the past year. Second, we will identify the duration of residence at the time of the extreme weather events. Third, we will estimate the association between the extreme weather events and the stress level of individuals. Fourth, we will estimate the mediation effect of the duration of residence on the association between the extreme weather events and the stress level of individuals.

Primary and secondary outcomes

The primary outcome is a derived stress score from the questionnaire for self-reported stress level in the health and exposure survey. This score will reflect the immediate stress level of the participants. The secondary outcomes are self-reported medication records where mental illness medications are identified. Since a few participants had reported the medication records, the sample size will be decreased. The individuals who experienced the number of extreme weather events in various time frames including one to three months, three to six months, six to twelve months, and in the past year will be compared.

Description and rationale for desired data from PEGS (survey data, demographic data, genetic data, etc.)

We request following variables for the analysis:

- Demographic data: basic information of the participants including sex, race/ethnicity, study events, coordinates
- Survey data
  - Health & Exposure survey
  - Exposome survey (internal & external)
    - ATC codes from self-reported medication
- Genetic data
  1. FKBP5 gene: SNPs – rs1360780, rs3800373, rs9296158
  2. CRHR1 gene: SNPs – rs7209436, rs110402, rs242924
  3. BDNF gene: SNP – rs6265 (Val66Met)
  4. SLC6A4 gene: 5-HTTLPR
  5. COMT gene: SNP – rs4680 (Val158Met)
  6. NR3C1 gene: SNPs – rs6198, rs41423247 (Bcl1 polymorphism), rs5522
  7. SERPINA6/SERPINA1 gene: SNPs – rs941601, rs8022616
- Residential & Neighborhood contexts
  - Environmental Justice Index variables
  - Socioeconomic Vulnerability Index variables

– Earth Observation variables

Description of desired populations (gender, race/ethnicity, age, genotype, phenotype, etc.)

Since we are interested in the general population, we do not have any specific criteria for the desired populations. We will include all participants who have completed the health and exposure survey and have the residential history data. For sensitivity analysis, we will request genotypes (single nucleotide polymorphisms) for controlling confounders. We expect that the sample size with full covariate data are more than 1,000.

Details of study design

The main outcome is assessed from the questionnaire in the health and exposure survey.

Main exposure is measured from the NOAA Storm Events Database. The database records twelve types of weather events including tornado, thunderstorm wind, hail, flash flood, flood, winter storm, ice storm, snow, extreme cold, extreme heat, drought, and high wind. The database provides the date, time, location, and intensity of the events. The intensity of the events is measured by the number of fatalities, injuries, and property damages. Locations and the intensity of the events will be used as the main exposure.

The duration of residence will be measured by the time difference between the date of the extreme weather event and the date of the residential history. The stress level of individuals will be measured by the stress score derived from the questionnaire in the health and exposure survey.

Statistical analysis plan including power analysis

G-computation (or G-formula) models are used to examine the effect modification of the cumulative exposure to extreme weather events assessed from NOAA Storm Events Database. Stress scores will be considered ordinal and continuous variable, each of which will be analyzed with multinomial logistic and quantile regression models, respectively. These models will serve as a base model for g-computation. Treatment regime is derived from the cumulative count of EWEs in fixed time windows (e.g., 5- and 10-year).

Any other relevant supporting data

We will operationalize the exposure to extreme weather events using NOAA Storm Events Database. Spatial covariates measured at residential locations

such as greenspace and bluespace will be derived from the National Land Cover Database, the National Hydrography Dataset, and USGS Protected Area Database.