Quantifying the Association Between Environmental Quality and Mental Disorders

Atif Khan[1], Hannach Lendecker[2], Patrick Sullivan[3], and Andrey Rzhetsky[1 4]

[1]Department of Medicine, Section of Genetic Medicine, University of Chicago, USA
[2]The Institute for Society and Genetics, The University of California, Los Angeles
[3]Karolinska Institute and Sweden, UNC Hospitals - Chapel Hill
[4]Conte Center for Neuropsychiatric Genomics, Institute for Genomics and Systems Biology
atifkhan@uchicago.edu, andrey.rzhetsky@uchicago.edu

THE UNIVERSITY OF CHICAGO DEPARTMENT OF MEDICINE

1. Research Objectives

- This study reports computational investigation for the etiology of mental disorder by using private health insurance claims (EHR data). A comparative analysis of following six mental disorders is presented in this study:
  1. Bipolar Disorder
  2. Schizophrenia
  3. Personality Disorder
  4. Major Depressive Disorder
  5. Epilepsy
  6. Parkinson's Disease
- From a comprehensive twin meta-analysis[1], the estimates for environmental contribution are 66% for major depression, 32% for bipolar disorder, and 23% for schizophrenia. The objective of this research is to:
  - Review, identify, and define the potential environmental risk factors in the context of mental disorders.
  - Build a population level computational model to investigate the disease-environment association by utilizing data from serval public and private data sources.
  - Finding the putative contributors to the disease along with the plausible biological mechanisms to better understand the disease.

2. Material and Methods

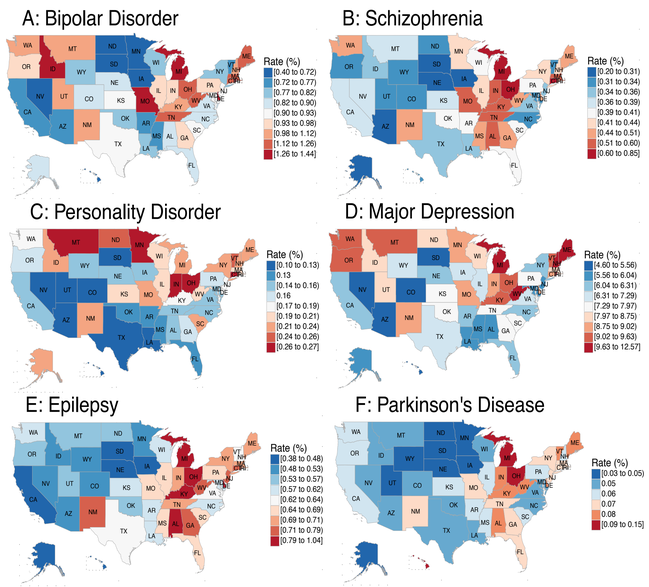
- The dataset includes time-stamped patient treatment episodes with individual patient diagnoses. The disease count was defined as patients with a specific diagnostic claim over the period of 2003-2013 identified in MarketScan database by the corresponding International Classification of Diseases version 9 (ICD-9) code.
- We examined the health insurance claims record of 151,104,811 unique U.S. individuals with the total unique cases of Bipolar Disorder: 1,238,255, Schizophrenia: 831,426, Personality Disorder: 234,341, Major Depression: 10,041,023, Epilepsy: 937,933, Parkinson's Disease: 243,493.
- We used a mixed-effects regression, modeling disease counts per age and sex group per county with a Poisson distribution, where the rate of the Poisson distribution is defined by a logarithm of a linear combination of predictors and random effects.
- We implemented this approach using Monte Carlo Markov chain (MCMC) algorithms. We modeled these variables comparing different septiles to estimate relative prevalence ratios (PR) and credible intervals (CI).
- County and state level random effects model the unknown factors that vary geographically and govern differences in county-specific disease rates after accounting for all fixed effects.
- The county-level predictor variables used in this model were obtained from the following databases. All predictors were transformed into septiles (7 groups Q1 to Q7) where higher septile represent higher numbers and percentages

Category	Names	Description	Data Sources
Disease Counts	Bipolar Disorder, Depression, Schizophrenia, Personality Disorder, Parkinson's Disease, Epilepsy.	The disease counts were obtained from the health insurance claims data (diagnostic and prescription histories) representing over 150 million unique U.S. patients. The database was curated from more than 100 insurance carriers and large self-insuring companies for the years 2003 to 2013.	MarketScan Health Research Database[1]
Environmental Quality	Air Quality	Quality index based on the concentrations of 87 hazardous air pollutants (toxics) obtained from 7 different data sources.	U.S. EPA[2]
	Water Quality	Quality index based on the 80 indicators of water quality obtained from 5 different data sources.	U.S. EPA
	Land Quality	Quality index based on 26 indicators obtained from 11 data sources of agricultural, industrial facilities, geology/mining and land cover.	U.S. EPA
	Built Quality	Quality index based on the 14 indicators form 4 different data sources consisting of traffic-related data, transit access, pedestrian safety, access to various business environments (such as the food, recreation, health care, and educational environments), and the presence of subsidized housing.	U.S. EPA
Weather Quality	Weather Quality	Represented by the average number of "good days" and "bad days" during 2003-2010 for each county. These averaged measures were obtained from daily weather characteristics.	NLDAS[3], DAYMET[4]
	Daily Weather	Consists of daily temperature (minimum and maximum), wind, specific humidity, solar radiations, cloud index, and precipitation.	NCEI[5], DAYMET
Socio-economics/ Socio-demographics	Urban/Rural	Percentage of rural and urban population within each county.	U.S. Census[6]
	Population Density	Number of people per square mile of area.	U.S. Census
	Race/Ethnicity	The percentages of racial composition of the county.	U.S. Census
	Health Insurance	Percentage of insured population.	U.S. Census
	Economy	Percentage of poor population.	U.S. Census

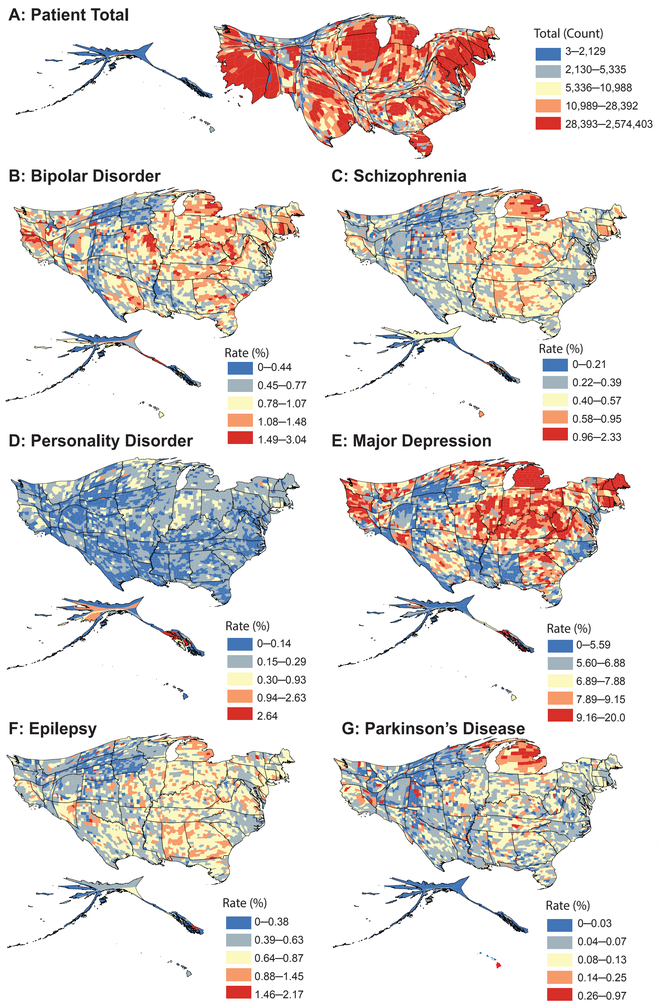
1:Truven Health Analytics MarketScan health claims database, 2:Environmental Protection Agency (EPA), 3:North American Land Data Assimilation Systems (NLDAS), 4:Daily Surface Weather and Climatological Summaries (DAYMET), 5:National Centers for Environmental Information (NCEI), 6:United States Bureau of the Census.

3. Estimated Prevalence

- The estimated prevalence of treated bipolar disorder is 0.82%, schizophrenia: 0.55%, personality disorder: 0.15%, Parkinson's disease: 0.16%, epilepsy: 0.62% and the major depression: 6.64%.

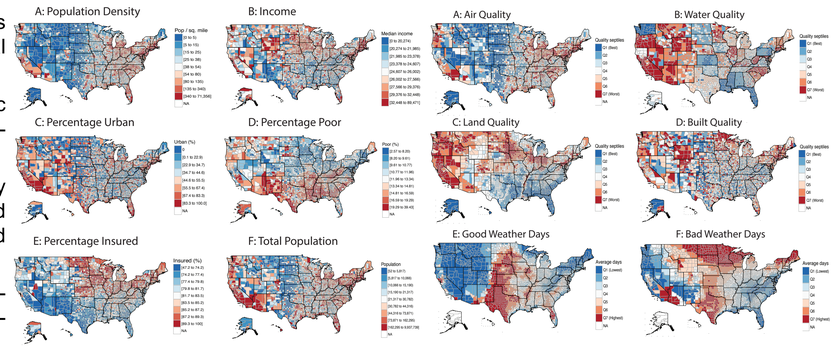


- The county level prevalence of many disease shows several strong geo-spatial clusters.



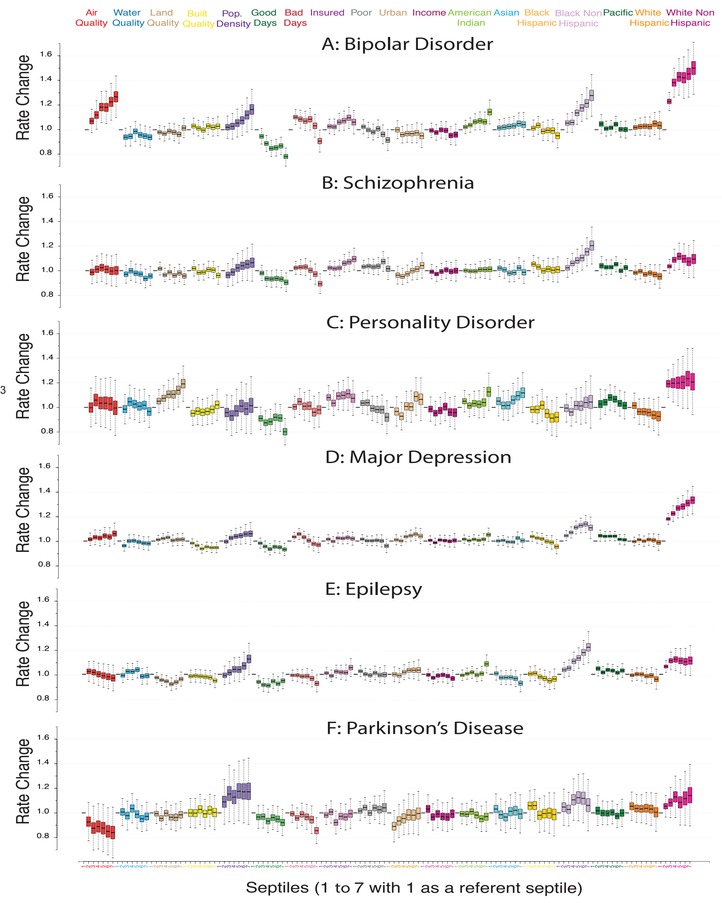
4. Environmental and Sociodemographic Factors

- Data obtained from 2010 U.S. Census was used to link the county-level percent racial distributions for the following groups:
  - American Indian - Asian - Black Hispanic - Black non-Hispanic - Pacific Islander - White Hispanic - White non-Hispanic
- The domain specific environmental quality indices for air, water, land, built, and sociodemographic qualities were obtained from U.S. EPA.
  - Air: 89 variables - Water: 80 variables - Land: 26 variables - Built: 14 variables - Weather: 8 variables



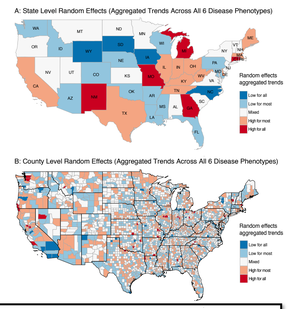
5. Results

- Estimated rate of disease from mixed effect Poisson regression model.



6. Random Effects

- Random effects are measured at the State and the County levels showing dissimilar distribution across all the mental conditions.
- State-level random effects likely absorb state-specific differences in both disease reporting and the true prevalence.
- County-level random effects can be thought of as residual variation not explained by our fixed-effect predictors and state-level random effects.



7. Conclusion

- We found that the spatial patterns of apparent prevalence of six mental disorders remarkably different in the US.
- The US East coast experiences higher prevalence of these phenotypes than the West coast.
- Geo-spatial clusters with high prevalence of major depression are observed among almost all the counties of Michigan, New Hampshire and Maine.
- The strongest predictor of mood disorders (major depression and bipolar disorder) in a county is its percentage of White individuals. By contrast, a higher percentage of Black, non-Hispanic individuals is associated with higher rates of schizophrenia and epilepsy.
- The strongest predictor for bipolar disorder, next after ethnic composition of population, is air quality: polluted air is associated with an approximate 27% increase (95% CI: [15%, 40%], $p < e^{-4}$) in the apparent rate of bipolar disorder.
- The counties with the highest number of pleasant weather days (Q7) are associated with an estimated 22% (95% CI: [17%, 27%], $p < e^{-4}$) decrease in the rate of bipolar disorder.
- The validation of the model by considering the individuals with at least 2 separate visits with diagnosis (ICD9 codes) showed similar trends and comparable rates.

Plausible Biological Mechanisms

We found at least two animal studies that provide concrete experimental evidence of the connection between air pollution and psychiatric disease.

1. Healthy dogs exposed to traffic-related pollution in Mexico City, followed by detailed analysis of tissue pathology and physiological effects in canines[2]. This study suggests three possible pathways:
   - Indirect, due to inflammation in lungs, leading to systemic inflammation.
   - Direct, transport to brain by intravascular brain macrophages.
   - Direct, transport of pollutants to brain via nasal respiratory damage.
2. Healthy 4-week-old male mice were exposed to urban-like to ambient fine airborne particulate matter pollution in laboratory conditions for 10 months[3].
   - Analysis of tissues of the exposed animals revealed inflammation of brain tissues, especially in hippocampus.
   - Exposed animals showed signs of cognitive impairment in spatial learning and memory, and depression-like behavioral symptoms.

References

1. Polderman, Tinca JC, et al. "Meta-analysis of the heritability of human traits based on fifty years of twin studies." *Nature genetics* 47.7 (2015): 702.
2. Calderon-Garciduenas, Lilian, et al. "Air pollution and brain damage." *Toxicologic pathology* 30.3 (2002): 373-389..
3. Fonken, Laura K., et al. "Air pollution impairs cognition, provokes depressive-like behaviors and alters hippocampal cytokine expression and morphology." *Molecular psychiatry* 16.10 (2011): 987.

Acknowledgements

We would like to thank Chris Lyttle and Atul Kumar for their support in the obtaining and pre-processing the data.