Supplementary Online Content


**eMethods.** Weather Data Sources, Clinical Data Sources, Outcome Measures, and Statistical Analysis

This supplementary material has been provided by the authors to give readers additional information about their work.
Weather Data Sources

Thunderstorm events were identified using the U.S. National Lightning Detection Network (NLDN) database maintained by the U.S. National Oceanic and Atmospheric Administration (NOAA). NLDN monitors lightning activity across the U.S. in real-time through a network of lightning sensors that use magnetic direction and time-of-arrival measurements to accurately identify each lightning event. In the contiguous United States, NLDN is estimated to capture more than 95% of cloud-to-ground lightning flashes (only cloud-to-ground lightning flashes were included in this study; cloud-to-cloud lightning was not included).\(^1\) Our analysis used aggregate lightning data published by NOAA that provided counts of all lightning strikes within a specific county and day from January 1999 to December 2012.

In the primary analysis, thunderstorms were identified at the county-day level based on the presence of any cloud-to-ground lightning flashes and positive precipitation. In additional analyses, alternative definitions of thunderstorms were used to capture storms of varying severity, including county-day observations with any lightning flashes (with or without precipitation), as well as county-day observations with lightning, precipitation, and above-median wind speed on the lightning day. The dataset contained approximately 16 million county-day observations (3,127 counties each observed on approximately 5,110 days).

Temperature and precipitation data were obtained from NOAA’s Global Historical Climatology Network (GHCN) station monitoring data. Wind speed was obtained from NOAA’s Global Summary of the Day (GSOD) station monitoring data and North American Regional Reanalysis (NARR) data. Data on air pollution, including concentrations for fine particulate matter (PM\(_{2.5}\)), coarse particulate matter (PM\(_{10}\)), ozone (O\(_3\)), nitrogen dioxide (NO\(_2\)), sulfur dioxide
(SO₂), and carbon monoxide (CO), were obtained from the U.S. Environmental Protection Agency’s Air Quality System (AQS) station monitoring data. Daily total pollen count data were obtained from the American Academy of Allergy Asthma & Immunology (AAAAI) and were available for 61 monitoring sites in 59 counties.

County-day level averages were computed for precipitation, wind speed, temperature, and air quality using station-monitoring data. Because monitoring station networks are often sparse, not all counties have weather or pollutant monitoring stations within their geographic boundaries. To maximize sample coverage, we created county-level weather and pollution measures using an inverse distance weighting approach. Specifically, for each county, we identified monitoring stations that were within 20 miles of the county’s geographic centroid and then computed mean weather and pollution measures across these stations, weighting each station’s observations by the inverse of the station’s distance to the county’s centroid.

Clinical Data Sources

Health care utilization and demographic data on Medicare beneficiaries aged ≥ 65 years were obtained from the Medicare Provider Analysis and Review (MedPAR) file, the Outpatient Standard Analytic Files, and the Master Beneficiary Summary File (MBSF, including the Chronic Conditions segment). Data files covered Medicare claims and enrollment information for 100% of beneficiaries during the study period, 1999-2012. MedPAR was used to identify all emergency room visits for respiratory illness that resulted in hospitalization and the outpatient file to identify visits that did not result in hospitalization. History of asthma or COPD was obtained from the Chronic Conditions segment of the MBSF. For each calendar day, we identified all living fee-for-service (FFS) beneficiaries whose first diagnosis of either asthma or COPD occurred on or before
that day. To ensure accurate measures of chronic conditions, analyses were restricted to beneficiaries continuously enrolled in fee-for-service Medicare for 2 years.

**Outcome measures**

Our primary outcome was the county-level rate of daily emergency department (ED) visits by any Medicare beneficiary with a primary diagnosis of any respiratory illness as defined by *International Classification of Diseases, 9th Revision* (ICD-9) codes 460-519. A broad definition for respiratory illness was used since specific categories such as asthma, COPD, or reactive airways disease may be unrecognized, undocumented, or not coded as the primary diagnoses by the treating physician. The fraction of Medicare beneficiaries in a county who had an ED claim for respiratory illness was computed for each day, both overall and for beneficiaries with a history of asthma or COPD.

**Statistical Analysis**

The relationship between thunderstorm events, climatological and air pollutant changes, and respiratory illness was estimated using an event-study approach that focused on changes in environmental and health outcomes in the days before versus after a given thunderstorm event. In the event-study analysis, we calculated the differences in outcomes compared to the storm day (day zero) for each day in the selected period before or after a storm. We then aggregated the data across all storms in our sample allowing study and visualization of average changes in the days surrounding storms. This analysis was premised on the assumption that the specific timing of thunderstorm events, conditional on covariates such as seasonal indicator variables, was uncorrelated with non-thunderstorm-mediated factors that may also affect respiratory illness, a quasi-experimental approach. The following county-day-level statistical model was estimated for
each of several dependent environmental variables (precipitation, wind speed, temperature, fine particulate matter, and pollen) as well rates of ED visits for respiratory illness:

\[ Y_{ct} = \alpha + \sum_{j=-19}^{20} \beta_j \text{Thunderstorm}_{c(t+j)} + \text{Covariates}_{ct} \quad (1) \]

where \( Y_{ct} \) was the outcome for county \( c \) on day \( t \); \( \text{Thunderstorm}_{c(t+j)} \) was a series of binary indicators for whether a thunderstorm occurred in county \( c \) on the \( j \)th day since day \( t \); \( \text{Covariates}_{ct} \) denote fixed effect indicators for county (approximately 3,100 indicators), year (14 indicators), month-of-year (12 indicators), and day-of-week (7 indicators). Results of this regression were first presented by plotting the estimated values for \( \beta_{20}, \beta_{19}, \ldots, \beta_{1}, \beta_{0}, \beta_{-1}, \ldots, \beta_{-18}, \beta_{-19} \) and their associated 95% confidence intervals, tracing out the evolution of environmental and health outcomes 20 days before and 20 days after a thunderstorm event occurred. To facilitate interpretation overall and between sub-groups, the coefficients were also summed in the \( \pm 3 \) days surrounding thunderstorms to depict with a single point estimate the relationship between thunderstorm events and emergency visits for respiratory illness. Robust variance estimators were used in all analyses to account for clustering of outcomes and thunderstorm events within counties.

The relationship between emergency respiratory visits and thunderstorm events was estimated for Medicare beneficiaries overall and separately for beneficiaries with a prior history of asthma or COPD. In addition, in sub-group analyses with formal tests for interactions, thunderstorms were stratified by severity into three categories (any lightning, lightning with precipitation, and lightning with precipitation and high winds), an analysis conducted to assess whether increases in emergency respiratory visits were greater for more severe storms. Finally, because thunderstorms could in theory lead to increases in medical care for reasons not specifically related to respiratory health, a falsification analysis was conducted using ED visits for three
conditions unlikely to be affected by thunderstorms: sepsis (ICD-9: 995.91), pulmonary embolism (ICD-9: 415.1), and deep vein thrombosis (ICD-9: 453.4).³

Analysis were performed in Stata (v. 14). The 95% confidence interval around reported estimates reflects 0.025 in each tail or \( P \leq 0.05 \). The study was approved by the institutional review board at the National Bureau of Economic Research. The Medicare data used were part of a larger NIH project to study the impact of environment on health using econometric methods such as event study analyses and natural experiments. The data use agreements for that project included Medicare data up to 2012 and not beyond; as a result, the years analyzed in the current study were restricted to 2012 and earlier.
REFERENCES: