American Journal of Epidemiology
© The Author(s) 2017. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
Vol. 186, No. 11
DOI: 10.1093/aje/kwx194
Advance Access publication: June 2, 2017

Practice of Epidemiology

Measuring the Impact of Disasters Using Publicly Available Data: Application to Hurricane Sandy (2012)

Steven J. Mongin*, Sherry L. Baron, Rebecca M. Schwartz, Bian Liu, Emanuela Taioli, and Hyun Kim

* Correspondence to Steven J. Mongin, Division of Biostatistics, School of Public Health, University of Minnesota–Twin Cities, 2221 University Avenue SE, Suite 200, Mail Code MMC2702A, Minneapolis, MN 55414-3075 (e-mail: sjmongin@umn.edu).

Initially submitted August 19, 2016; accepted for publication February 8, 2017.

The unexpected nature of disasters leaves little time or resources for organized health surveillance of the affected population, and even less for those who are unaffected. An ideal epidemiologic study would monitor both groups equally well, but would typically be decided against as infeasible or costly. Exposure and health outcome data at the level of the individual can be difficult to obtain. Despite these challenges, the health effects of a disaster can be approximated. Approaches include 1) the use of publicly available exposure data in geographic detail, 2) health outcomes data—collected before, during, and after the event, and 3) statistical modeling designed to compare the observed frequency of health outcomes with the counterfactual frequency hidden by the disaster itself. We applied these strategies to Hurricane Sandy, which struck the northeastern United States in October 2012. Hospital admissions data from the state of New York with information on primary payer as well as patient demographic characteristics were analyzed. To illustrate the method, we present multivariate logistic regression results for the first 2 months after the hurricane. Inferential implications of admissions data on nearly the entire target population in the wake of a disaster are discussed.

counterfactual inference; disasters; finite population; Hurricane Sandy; public data

Abbreviations: CI, confidence interval; DRG, Diagnosis-Related Group; FEMA, Federal Emergency Management Agency; HCUP, Healthcare Cost and Utilization Project; MDC, Major Diagnostic Category; PPC, Primary Payer Class; ZCTA, Zip Code Tabulation Area.

Whether natural or man-made, the toll of a disaster on human health can be wide-ranging, persistent, and difficult to detect. Postdisaster behavior can create unanticipated consequences, compounding the harm—such as the increased incidence of carbon monoxide poisoning after widespread flooding due to the use of portable generators, pumps, or other appliances intended for outdoor use (1). Major floods can impact mental health, with manifestations in the form of posttraumatic stress disorder, anxiety, and substance abuse (2–5). The Centers for Disease Control and Prevention offers guidance on postdisaster surveillance (6) for ascertaining the acute needs of affected communities, including recommendations on survey sampling and inferential techniques aided by census tract data. While valuable, such investigations call for fieldwork, which may require considerable resources.

To overcome these challenges, researchers access large databases of hospital records, including information on inpatient care and emergency department visits (7–11). These data may be rich in detail but are often limited to host agencies or institutions, require a lengthy application process for acquisition, or are unique to particular states. One of our aims in this article is to describe publicly available data resources with sufficient detail on disaster exposure, and on health-care utilization both before and after the event, to facilitate these investigations.

A second aim is to consider the measurement of a disaster's impact on health. A general approach is to identify unexpected occurrences of a specific disease within a defined postdisaster window-period using predisaster rates as a reference (7, 8). It is also common to identify postdisaster disease patterns without prior specification of a disease (10). In other contexts, hospital

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

encounter-level data from the Healthcare Cost and Utilization Project (HCUP) (12) have been accessed for assessment of outcomes and trends (13–15). Location can also play an important role in evaluating the health impact of a disaster. While data resources for disaster exposure, such as evacuation zones and floodwater estimates, are publicly available, they rarely appear in disaster research (11).

Secular health trends appear in some of these studies but are often unaccounted for in the estimation of the effects of a sudden disaster, as in the comparison of hospital admission rates. Contrasts between pre- and postdisaster rates are made, but ideally we would compare the postdisaster rate with the counterfactual *nondisaster* rate—that is, the admission rate that would have occurred in the absence of the disaster (16). Long-term trends cause the mean admission rate in a predisaster window of 1 year (for example) to miss this ideal reference rate, inducing bias in the comparison between postdisaster and reference rates. This is illustrated in Figure 1 by HCUP admissions data from the 12 months before Hurricane Sandy. There is a generally increasing pattern in diagnosis proportions for infectious/parasitic diseases, such that a 12-month average would likely underestimate the counterfactual proportion in November 2012, the first post-Sandy month. Comparison with the September proportion alone ignores information about variance and seasonal patterns, as well as trend, and would probably produce a similar bias.

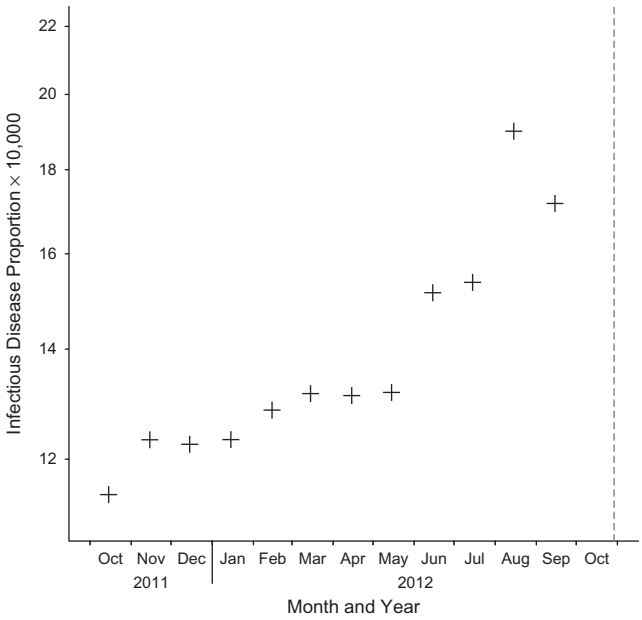

Figure 1. Proportions of infectious disease cases prior to Hurricane Sandy among hospitalized Medicaid patients residing on Long Island, New York, 2011–2012. Cases were identified through monthly hospital admissions data from the Healthcare Cost and Utilization Project (12). For each month, the mean daily probability of admission for the Major Diagnostic Category "infectious/parasitic" (infectious and parasitic diseases, systemic or unspecified sites) is indicated by a plus sign. The overall mean number of admitted patients per month (combining infectious disease and noninfectious disease) was 2,186 (standard deviation, 119). The vertical dashed line marks Hurricane Sandy landfall on October 29, 2012. The logit scale is used along the ordinate (*y*-axis) of the plot, with tick-mark labels scaled up by a factor of 10,000.

In this article, we use a simple statistical model that includes an interrupted time series (17, 18) to illustrate how bias due to time trends or seasonal patterns can be reduced, and we extend this model to compare health effects by geographic region. We further describe how publicly available, low-cost data on hurricane damage and hospital admissions can be combined with the appropriate statistical analysis to estimate the health effects of a disaster, with an example of application to Hurricane Sandy and Long Island, New York.

METHODS

The target population was defined as all residents of Long Island, New York, which contains Nassau and Suffolk counties, in October 2012. Data from the US Census Bureau (19) were used to describe this population by age, sex, race, and Hispanic ethnicity. A geospatial map provided a picture of the distribution of residents throughout the region.

Residential exposure to Hurricane Sandy

Data on exposure were obtained from the Federal Emergency Management Agency (FEMA) via publicly available web sites created by FEMA's Modeling Task Force (20). This group integrates hazard and modeling information from several sources, including FEMA individual assistance inspectors, who are dispatched to damaged sites following a hurricane (21). For each of a series of dates throughout November and early December of 2012, the extents of damage and flooding were reported in geospatial data sets with corresponding information on latitude/longitude and Zip Code Tabulation Area (ZCTA). From these data, we defined 4 exposure measures:

1. *Damage severity*: For all reported sites post-Sandy, 5 levels of damage were assigned (range, 1–5; 1 = "no damage," 5 = worst damage, "destroyed"). The sum (across all sites) of the maximum of these values at each site was standardized by the population count, yielding a ZCTA-level measure of the amount of damage per resident.
2. *Damage site count*: The number of all distinct damaged or flooded sites within each ZCTA, standardized per 100,000 residents to make the categories easier to interpret (≤1, >1–100, >100–2,000, or >2,000).
3. *Flood depth*: Similar in structure to "damage severity." The sum (across all sites) of the maximum reported floodwater depth at each site, standardized by ZCTA.
4. *Flood coverage*: FEMA impact assessment included an indicator of flood coverage (0 = not inundated, 1 = inundated) at each site. The sum (across all sites) of the maximum reported indicator at each site was standardized to produce a ZCTA-level measure of the extent of water inundation per resident.

For exposure measures 1, 3, and 4, histograms of the standardized (per resident) measures for all 177 residential ZCTAs of Long Island were examined for natural cutpoints, resulting in 4 levels: low, moderate, substantial, and high (or "severe" for damage severity). A geospatial map of each measure illustrated the distribution of exposure levels by ZCTA.

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

Health outcome measures

To capture the most acute effects of Hurricane Sandy, we determined health outcomes according to hospital admissions via the emergency department. Monthly inpatient data from January 2011 through December 2013 for Long Island residents were obtained from HCUP (12), whose State Inpatient Databases are sponsored by the Agency for Healthcare Research and Quality and are claimed to "encompass more than 95% of all U.S. hospital discharges" (22). Each record included patient demographic characteristics (residential zip code, age, sex, race, Hispanic ethnicity), admission month and hour, primary diagnosis by Major Diagnostic Category (MDC; version 24) and Diagnosis-Related Group (DRG; version 24) (23), vital status, and primary payer. Precise admission dates are withheld by the state of New York to protect patient privacy. By aggregating subsets of primary payers, we defined 4 Primary Payer Classes (PPCs): Medicare, Medicaid, private, and other.

HCUP employs a number of data quality control procedures (24); however, some additional data processing was warranted. New York is one of approximately 12 states (this number varies by year) that provide a "visit linkage" variable for tracking individual patients over time. Visit linkage information was available for 99.3% of all admission records; the remaining 0.7% of admissions were assumed to represent separate patients. For any patient with inconsistent race or Hispanic ethnicity reported across successive visits, the most commonly observed nonwhite race reported for that patient was imputed for all of their records. Admission records following reported deaths were eliminated, as were duplicate admission records, which were defined by having a common patient identifier, zip code, admission month and hour, MDC, and PPC. The MDCs comprise 25 health outcomes, ranging from "diseases and disorders of the nervous system" to "human immunodeficiency virus infections," though information in the latter category is withheld by New York State to protect patient privacy. Finer diagnostic information is given through the DRG code.

Statistical analysis

A hierarchical mixed-effects logistic regression model was used to estimate the daily probability of encountering an MDC patient among all admitted patients, based on the mean proportion of days per month that each MDC patient was admitted. For the ith patient in the jth month, the logit of this proportion (p_{ij}) was assumed to have the following structure:

$$\text{logit}(p_{ij}) = \beta_0 + \beta_{0i} + \beta_1 t_j + \beta_{2,k} b(k, t_j) + \beta_3 d(t_j) + (\text{demographic effects}), \quad (1)$$

with parameters to be estimated including

β_0, β_{0i} = intercepts, overall (fixed) and patient-specific (random), respectively;

β_1 = effect of continuous time (t_j), with t_j centered on the jth month;

$\beta_{2,k}$ = effect of the kth bimonthly pair (January–February, March–April, etc.; $k = 1$–6), for seasonal impacts, with bimonth indicator $b(k, t_j)$; and

β_3 = effect of the postdisaster month(s) of interest (e.g., November 2012), with indicator of postdisaster month(s) $d(t_j)$.

Demographic effects consisted of age (and its square), sex, race, Hispanic ethnicity, and age-sex interaction. In a given month, a patient could be admitted more than once for the same MDC. Days at risk of admission were reduced by half for concurrent deaths. The binomial response data for each patient-month combination was defined by the pair (number of admissions, number of days at risk). For example, if a patient was admitted twice with the same MDC in November and later died during that month, then their response pair was (2, 15). Another, more common example is a patient who was admitted in November with a non-MDC diagnosis and survived the month, producing the response pair (0, 30). For health effect measurement, we are primarily interested in β_3, which represents the postdisaster interruption in the time series (17, 18) of admission proportions established from predisaster information. Transformation of this parameter as $\exp(\beta_3)$ yields the odds ratio that compares the postdisaster MDC admission proportion with that projected from predisaster data. Regression analyses were restricted to those MDCs with at least 300 admissions overall.

Tier 1 (time and change-point analysis). Using the regression model of equation 1 for each MDC, we examined Hurricane Sandy effect phases from October 2012 to March 2013 by individual month and by successive pairs of months. For each effect phase and for each PPC and MDC, the ratio comparing the odds of encountering an admitted MDC patient with the pre-Sandy prediction of these odds was estimated by exponentiation of $\hat{\beta}_3$ (the fitted value of β_3). The model adjusts for temporal effects by: 1) continuous time (through $\beta_1 t_j$), tracking secular trends in admissions, and 2) a bimonthly series of 6 terms ($\beta_{2,1}, \ldots \beta_{2,6}$), to account for annually recurring effects. Combined with the demographic covariates, these temporal adjustments reduce bias due to long-term trend or seasonality, as compared with simple contrasts of pre- and post-Sandy data.

Tier 2 (time × region effect). Hurricane Sandy effect phases were similarly examined from October 2012 to March 2013. While the tier 1 analysis measures effects across Long Island overall, this extension measures the differential effects of damage and flooding by residential region. For every combination of effect phase, PPC, MDC, and each of the 4 residential exposure metrics (damage severity, etc.), a discrete, fixed covariate was added to the tier 1 model (equation 1) to represent the increasing levels of exposure by geographic region, as

$$\text{logit}(p_{ij}) = (\text{all equation 1 terms}) + \beta_{4,r}\, g(i, r) + \beta_{4,r}\, g(i, r)\, d(t_j), \quad (2)$$

where $\beta_{4,r}$ represents the effect of the rth region ($r = 1$–4), via exposure to damage or flooding, with geographic indicator $g(i, r)$. Inclusion of the main effect, $\beta_{4,r}\, g(i, r)$, prevents the confounding of hurricane damage effects with intrinsic regional differences.

Deviance residuals from the fitted logistic regression models were checked for evidence of lack of fit. Counts under 10 were censored from final tables per HCUP guidelines. All

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

statistics and graphics were produced in the R software environment, version 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria) (25). Geospatial data and graphics were processed using the "rgdal" package in R, version 1.1.7 (26). Mixed-effects logistic regression was performed with the "lme4" package, version 1.1.10 (27).

RESULTS

Demographic variables for Long Island as of 2012 are shown in Table 1. The population of nearly 3 million included over 400,000 residents (14.5%) at least 65 years of age and was 51.2% female, 79.3% white, and 15.6% Hispanic or Latino. A population density map for 2012 by ZCTA is displayed in Figure 2. Heavily populated areas in the southern regions of Long Island are evident, especially along the barrier islands where Hurricane Sandy approached. Choropleth maps of the 4 exposure measures illustrate their similarity in pattern, as shown in Figure 3. The southern parts of Long Island sustained the highest levels of damage and flooding. Some northern ZCTAs bordering Long Island Sound sustained severe effects, while most interior regions were spared. The measures of damage severity (measure 1) and damage site count (measure 2) incorporated wind damage.

Table 1. Demographic Characteristics of the Population of Long Island, New York, in 2012[a]

Characteristic	No. of Persons	%
Total population	2,837,087	100.0
Age, years		
0–19	740,356	26.1
20–44	872,305	30.7
45–64	813,386	28.7
65–84	349,076	12.3
≥85	61,964	2.2
Sex		
Female	1,452,322	51.2
Male	1,384,765	48.8
Race[b]		
White	2,248,512	79.3
Black or African-American	280,440	9.9
Asian	175,460	6.2
American Indian or Alaska Native	18,116	0.6
Native Hawaiian or other Pacific Islander	1,849	0.1
Other	172,693	6.1
Hispanic ethnicity		
Hispanic or Latino	441,444	15.6
Not indicated	2,395,643	84.4

[a] Population statistics for Long Island (Nassau and Suffolk counties); based on March 25, 2016, data from the US Census Bureau (19).

[b] Counts by race include multiple races reported by some survey participants. Each percentage is relative to the overall total population, so the percentage total for race may exceed 100.

Individual diagnostic categories

To illustrate the regression model, we show results for the first 2 post-Sandy months, November and December 2012. MDCs were examined separately by their fitted odds ratios. The volume of output motivated a combination of tabular and graphical information to convey the results of the tier 1 analyses for several MDCs, as shown in Figure 4. For each PPC, the MDCs with the top 7 odds ratio estimates (limited to clarify presentation) appear in decreasing order, along with numbers of hospital admissions before and after Hurricane Sandy. This is one of several similar displays that were produced as a single PDF file for Web Figures 1–11 (available at https://academic.oup.com/aje), allowing convenient review of results over Sandy effect phases from October to March, sequenced as October, October–November, November, November–December, etc. In the November–December phase shown, the "ear, nose, mouth, and throat" category appears in all 4 PPCs, and its frequency is significantly high among Medicare and Medicaid patients, where the estimated odds ratios are 1.59 (95% confidence interval (CI): 1.31, 1.93) and 1.43 (95% CI: 1.06, 1.93), respectively. In other words, among Medicare patients admitted to a hospital just after Hurricane Sandy, the odds of an "ear, nose, mouth, and throat" diagnosis were 59% higher than predicted for November–December 2012 by pre-Sandy data.

Numbers of respiratory admissions exceeded predictions for Medicare patients, with an odds ratio estimate of 1.11 (95% CI: 1.05, 1.18). Patients in the "other" payer class (including self-pay) experienced increased mental health admissions per the estimated odds ratio of 1.48 (95% CI: 1.11, 1.98). The example shown in Figure 1 used simple mean disease proportions for infectious/parasitic diseases among Medicaid patients. The corresponding regression model yielded an estimated odds ratio of 1.18 (95% CI: 0.94, 1.46), shown in Figure 4. Based on 21 months of data preceding Hurricane Sandy, the estimated slope for the tier 1 continuous time term is equivalent to an annual 18% ($P < 0.001$) increase in the odds of infectious/parasitic disease among admitted patients.

The tier 2 analysis provided estimates of the modifying effect of residency on the tier 1 effect, by comparing residual hospital admissions in differently affected regions of Long Island. Results are illustrated in Web Figure 12, which again covers the November–December effect phase. The odds ratios of interest compare higher regional flood coverage with the lowest level as the referent. Admission counts within regional exposure groups were small for some MDCs, yielding substantial apparent uncertainty. Despite this variability, the change in the number of admissions due to "digestive system" disorders was significantly higher in both Medicare and Medicaid payer classes, with estimated odds ratios of 1.20 (95% CI: 1.03, 1.39) and 1.32 (95% CI: 1.00, 1.75), respectively.

Analyzing by constituent DRGs

In addition to the MDC, each admission record in the HCUP data includes a more specific DRG (23). The full analysis of each MDC included estimation of the tier 1 effect for each of its associated DRGs, one of which is illustrated in Figure 5 for respiratory disorders. In the November–December time frame, the admission proportions of 4 DRGs were significantly higher

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

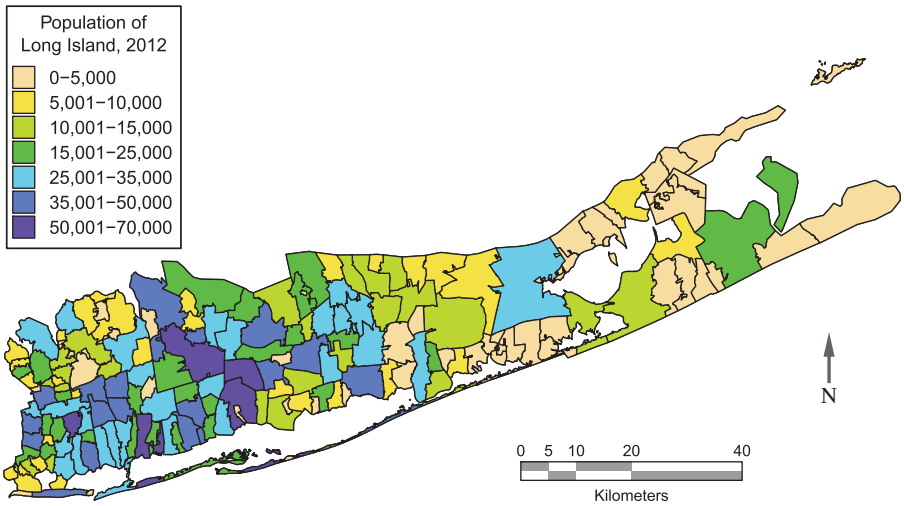

Figure 2. Distribution of the population of Long Island, New York, in 2012, by Zip Code Tabulation Area (19).

than predicted: 1) pneumothorax with complications and comorbidity; 2) bronchitis and asthma among persons aged >17 years with complications and comorbidity; 3) chronic obstructive pulmonary disease; and 4) respiratory system diagnosis with ventilator support <96 hours. The proportion of admissions for chronic obstructive pulmonary disease was significantly elevated in 3 of the 4 payer classes, and it was highest (though not significant) for the remaining "other" class.

DISCUSSION

Measuring health effects typically requires a baseline or denominator. Rather than define this baseline as a summary of predisaster data, we emphasize the value of projection from predisaster data to a postdisaster time frame. Due to the potential for regional health consequences, population-wide data are relevant for the study of disasters. In this article, we have described the accessing and processing of publicly available data for this purpose. Application to Hurricane Sandy provides a practical example of data utilization and statistical modeling designed to reduce bias.

Interpretation of results

In the tier 1 analysis, increases for "ear, nose, mouth, and throat" and respiratory outcomes in November and December may have resulted from the growth of mold in flooded residences and workplaces. The cited increase in mental health admissions in the "other" payer class is consistent with the study by Schwartz et al. (4, 5), who found significant associations between self-reported measures of hurricane exposure and perceived stress using a cross-sectional survey of volunteer participants. Elevations in digestive system disorders among the most flooded regions, as shown in the tier 2 analysis, were barely significant with 95%-level confidence intervals, but their separate appearance in both Medicare and Medicaid payer classes is suggestive. Extensive flooding may have exposed residents to contaminated water supplies or bacteria-laden flood residue. In our analysis of DRGs among privately insured patients admitted for respiratory disorders, the unexpectedly high number of admissions for ventilator support is similar to the finding of Lee et al. (11), who reported spikes in ventilator use in the first 2 days after landfall among New York City emergency departments.

Strengths and limitations

The publicly available data we obtained, with ZCTA-level precision on population, disaster exposure, and health outcomes spanning years, were useful resources for measuring the health effects of Hurricane Sandy. On Long Island, ZCTA populations range from a few hundred to over 50,000. A ZCTA with 100 damage reports among 500 residents poses greater risk than one with 100 reports among 50,000 residents. Standardization of the damage summaries by local population size accounts for this variability and thereby improves estimation of personal exposure to the hurricane. However, we emphasize that the health outcomes component does not require enumeration of the Long Island population at risk; it is based solely on HCUP admissions data.

Classes of health outcomes can be formed by aggregation of MDCs. Rather than excluding them, categories with low counts may be grouped by similarity; for example, in a separate analysis we combined "burns," an uncommon but potentially important postdisaster outcome, with the larger MDC "injuries/poisonings/toxic drugs." Opposite from aggregation, the finer DRGs can be used to examine more specialized endpoints, as shown for the respiratory category. The State Inpatient Databases translate disparate hospital records from 30 states into a uniform format, facilitating cross-state comparisons (22). Demographic and payer variables can be used to identify and compare health outcomes among subgroups.

The comprehensive nature of HCUP data provides an advantage for postdisaster, regional surveillance. If the inpatient data

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

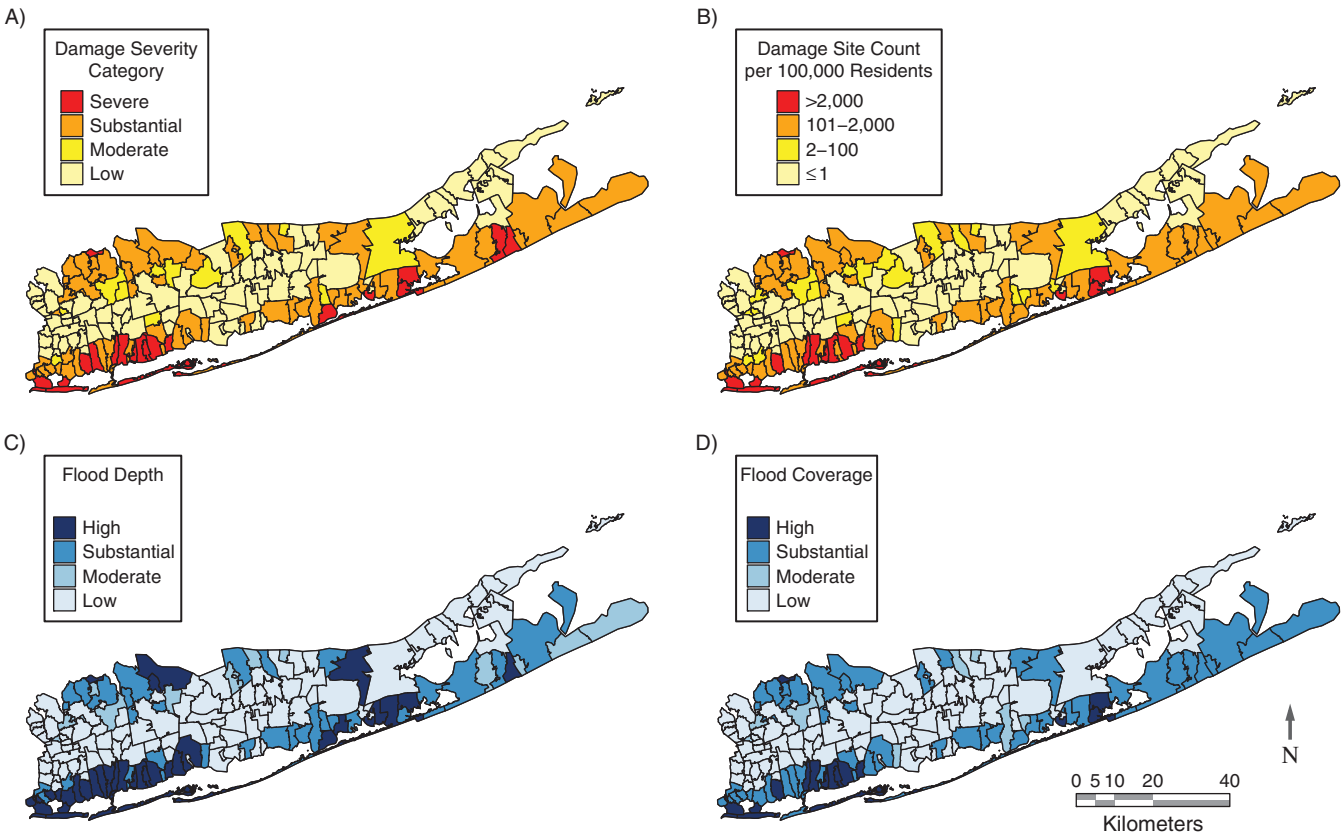

Figure 3. Extent of building/structural and flood damage on Long Island, New York, after Hurricane Sandy (October 29, 2012), by Zip Code Tabulation Area (ZCTA). Data were obtained from the Federal Emergency Management Agency (FEMA) (20, 21). A) Damage severity category, derived from 2015 FEMA data gathered following Hurricane Sandy. For each ZCTA, the damage severity score was based on the sum of all maximum damage assessments from each site within the ZCTA reported during the observation period of November and December 2012. For example, ZCTAs in the "severe" category typically contain some number of buildings or structures classified as destroyed or with major damage. B) Damage site count, based on the number of locations identified by FEMA to have some level of damage, regardless of severity. C) Flood depth, computed as the maximum of all site-specific water depths during the observation period. D) Flood coverage, derived from the count of all sites reported as "inundated." All metrics were normalized by ZCTA-level population counts.

indeed represent over 95% of all hospital admissions, we consider the implications for statistical inference. The defined target population is all residents of Long Island in October 2012, and the objective is post-Sandy surveillance for their health outcomes. Allowing for residents admitted to out-of-state hospitals, assume that HCUP data captured 80% of all admissions in the target population. Under this assumption, the standard confidence intervals presented in this article are substantially oversized, and correction to account for a finite population is warranted. An approximate correction factor (28) for the standard error is $\sqrt{1 - 0.8} \approx 0.45$, reducing confidence interval lengths by 55%.

The bias-reducing feature of the tier 1 model is exemplified in the result for infectious/parasitic diseases, where an odds ratio of 1.18 was found. Comparison of post- with predisaster odds over the previous year (for example), during which the fitted odds of diagnosis increased 18% on average, would yield a much higher ratio, likely with positive bias. In other words, the November–December proportion of infectious diseases was on track to rise, even *without* the hurricane. There are many methods and corresponding software tools with which to embellish the predictive component as desired by separate trend lines (by region, for example), nonlinear structure, Bayesian approaches, or autoregressive models (29, 30). Direct comparisons of regional diagnostic proportions would fail to account for preexisting differences between regions that are not controlled for by other model covariates. The tier 2 extension reduces this source of bias by comparing time-driven changes in diagnostic proportions among regions, essentially by means of a difference-in-differences approach. The tier 2 model can be modified to estimate differences among demographic subgroups as well.

There are several limitations in the application of this model to Hurricane Sandy that should be considered. In the days before landfall, many residents moved out of the most threatened areas, either on their own or with the assistance of caregivers. The effects of evacuation can be nontrivial, especially for the elderly, and such effects should be attributed to the hurricane. If any individual suffered physical injury or a mental health event that required emergency admission to a

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

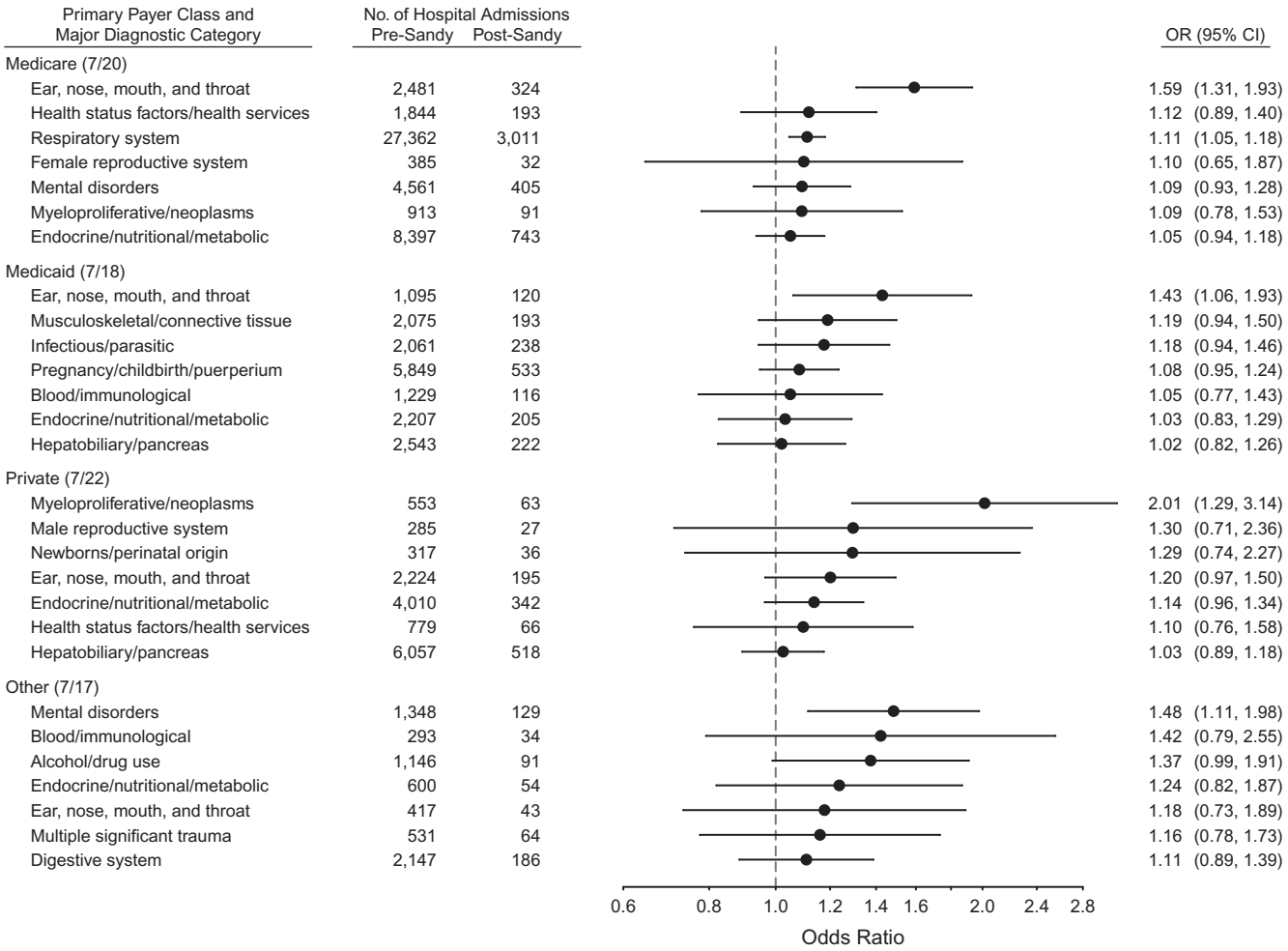

Figure 4. Numbers of pre- and post-Hurricane Sandy November–December hospitalizations and odds ratios (ORs) for emergency department admission by Primary Payer Class (PPC) and Major Diagnostic Category (MDC), Long Island, New York, 2011–2012. Admission counts and ORs, comparing the month of Sandy exposure with the odds projected from the pre-Sandy phase, were derived from a mixed-effects logistic regression model for binomial counts, accounting for repeated admissions within patients, and adjusting for age, sex, race, Hispanic ethnicity, and calendar time. In the plot, each point estimate of the OR is marked by a dot, with a corresponding horizontal line spanning a 95% confidence interval (CI). Within each PPC, based on decreasing estimates of the OR, the top 7 (of up to 23) MDCs are presented; the fraction in parentheses after each PPC name gives the number of MDCs displayed out of the number with a sufficient sample size for analysis. "Pre-Sandy" represents the period from January 2011 through September 2012; "post-Sandy" represents November through December 2012.

hospital outside of the state of New York, this event would not have appeared in the data available for this analysis. However, if they were admitted anywhere *within* the state, this admission was most likely captured. We concluded that this would cover many cases of evacuation as Hurricane Sandy approached. For situations where more coverage is desired, additional data could be obtained from participating states through HCUP.

The incidence of respiratory problems in the New York City area rises in November and remains elevated through the winter months (31). Many respiratory problems arising in the fall result in hospital admissions for treatment of asthma (31). Increased respiratory admissions in November and December were observed among Medicare patients in this study and may therefore be attributed to seasonal effects only. However, the role of the residual effects of the hurricane (such as accumulation of mold in waterlogged homes) cannot be completely discounted, since the model included calendar effects. Elevations observed in pre-Sandy months were adjusted for in the fitted models.

The State Inpatient Database for New York provided by HCUP is limited to calendar month for admission dates, has a reporting lag of 1–2 years, and is currently limited to 30 of the 50 US states. Our data required some processing but were otherwise consistent and well-documented. FEMA Modeling Task Force data were accessible and organized; however, they offer less documentation or support. The commitment of resources by FEMA may vary over time and by disaster. FEMA recently published an updated guide describing its commitment to playing a continuing role in both

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

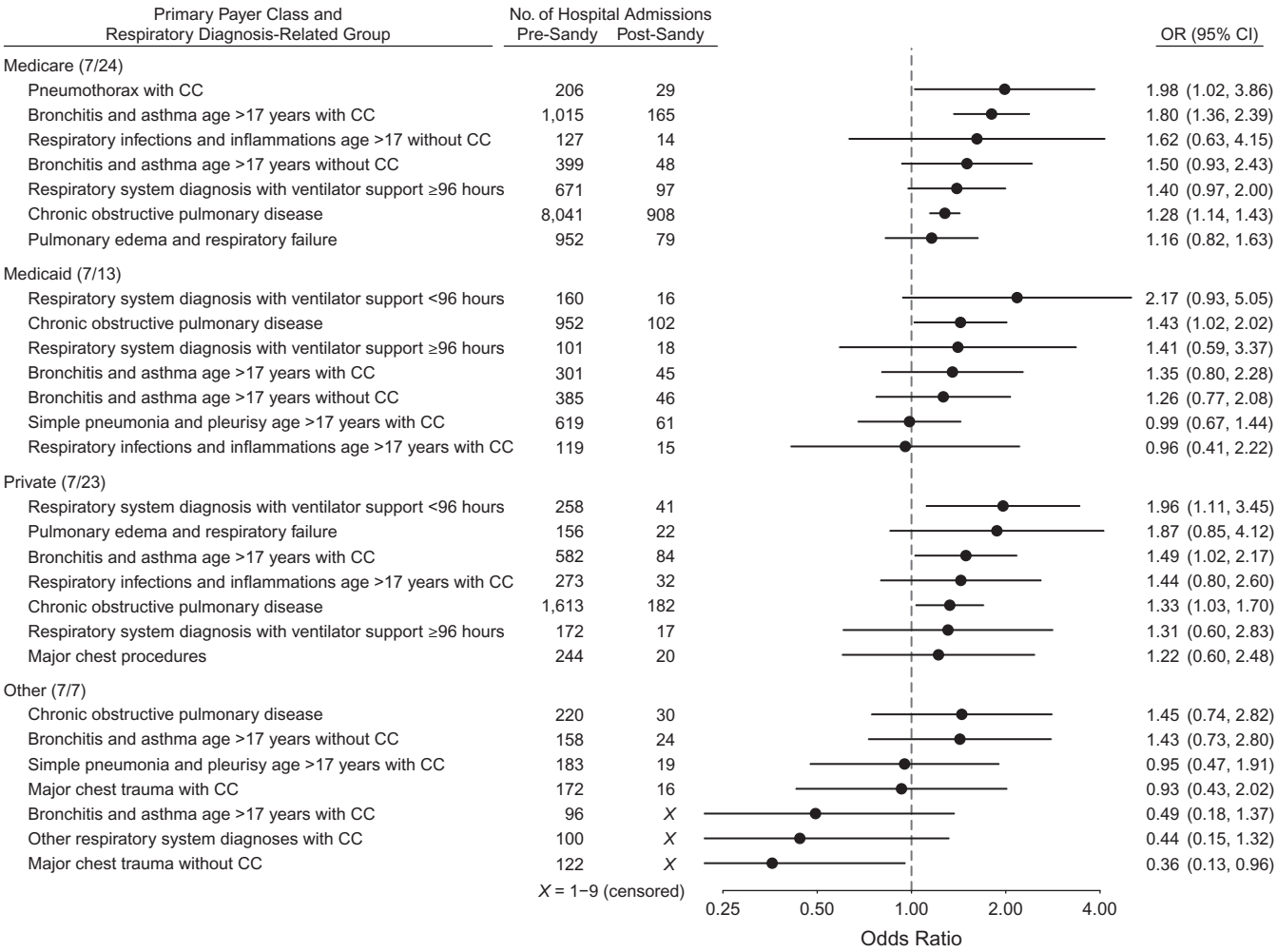

Figure 5. Numbers of pre- and post-Hurricane Sandy November–December hospitalizations and odds ratios (ORs) for emergency department admission by Primary Payer Class (PPC) and Diagnosis-Related Group (DRG) under the Major Diagnostic Category "respiratory system," Long Island, New York, 2011–2012. Admission counts and ORs, comparing the month of Sandy exposure with the odds projected from the pre-Sandy phase, were derived from a mixed-effects logistic regression model for binomial counts, accounting for repeated admissions within patients, and adjusting for age, sex, race, Hispanic ethnicity, and calendar time. In the plot, each point estimate of the OR is marked by a dot, with a corresponding horizontal line spanning a 95% confidence interval (CI). Within each PPC, based on decreasing estimates of the OR, the top 7 (of up to 29) respiratory system DRGs are presented; the fraction in parentheses after each PPC name gives the number of DRGs displayed out of the number with a sufficient sample size for analysis. "Pre-Sandy" represents the period from January 2011 through September 2012; "post-Sandy" represents November through December 2012. CC, complications and comorbidity.

individual and public assistance after a disaster, and the commensurate need for damage and loss assessment (32).

Our analysis of admissions data was based on proportions of admissions observed in each reporting period, which were negatively correlated across different disease categories. Acknowledging this caveat, compared with a count-based approach, admission proportions better insulate time-based effect estimates (tier 1) from secular changes in admission counts due to (for example) policy changes such as Medicaid expansion (33). For example, an increase in the number of infectious disease admissions in November may result from increases in all types of admissions, due to population-wide changes in health coverage. Furthermore, the distributions of these admission counts tend to depart from a natural Poisson distribution, due to a lack of counts exceeding 1 (underdispersion). Rare diagnoses are less affected by this issue, but for certain MDCs (e.g., respiratory disorders) and certainly for aggregated classes such as physical or mental health, this effect is substantial.

Conclusion

As a source of consistently gathered information over time, with nearly total population coverage, and with the person-specific linkage feature (in participating states), HCUP data are useful for estimating the effects of disasters, as shown in the tier 1 analyses. The level of detail is excellent along certain dimensions (diagnostics, administrative, patient linkage), but

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

less so on others (admission dates, reporting lag). For the retrospective review of admissions data, HCUP data are reasonably priced. For example, the 3 years' worth of State Inpatient Databases for New York that we obtained for this study were purchased for $1,200. We intend to make further use of this data source. The FEMA Modeling Task Force data were less user-friendly but still a valuable source for information on a variety of damage parameters and ZCTA-level accuracy. Comparison of postdisaster outcomes with regression-based *projections* from predisaster data—rather than predisaster summaries—is recommended for reduction of bias. With the careful use of these tools, publicly available data on exposure and hospital admissions can be used to measure the health impact of disasters.

ACKNOWLEDGMENTS

Author affiliations: Division of Environmental Health Sciences, School of Public Health, University of Minnesota–Twin Cities, Minneapolis, Minnesota (Steven J. Mongin, Hyun Kim); Barry Commoner Center for Health and the Environment, Queens College, City University of New York, New York, New York (Sherry L. Baron); Department of Occupational Medicine, Epidemiology and Prevention, Hofstra Northwell Health School of Medicine, Manhasset, New York (Rebecca M. Schwartz); and Department of Population Health Science and Policy and Institute for Translational Epidemiology, Icahn School of Medicine at Mount Sinai, New York, New York (Bian Liu, Emanuela Taioli). S.J.M. is currently with the Division of Biostatistics, School of Public Health, University of Minnesota–Twin Cites, Minneapolis, Minnesota.

This work was supported by the Office of the Assistant Secretary for Preparedness and Response (US Department of Health and Human Services) under grant HITEP150029-01-00.

Conflict of interest: none declared.

REFERENCES

1. Waite T, Murray V, Baker D. Carbon monoxide poisoning and flooding: changes in risk before, during and after flooding require appropriate public health interventions. *PLoS Curr*. 2014;6:ecurrents.dis.2b2eb9e15f9b982784938803584487f1.
2. Paranjothy S, Gallacher J, Amlot R, et al. Psychosocial impact of the summer 2007 floods in England. *BMC Public Health*. 2011;11:145.
3. Fernandez A, Black J, Jones M, et al. Flooding and mental health: a systematic mapping review. *PLoS One*. 2015;10(4):e0119929.
4. Schwartz RM, Sison C, Kerath SM, et al. The impact of Hurricane Sandy on the mental health of New York area residents. *Am J Disaster Med*. 2015;10(4):339–346.
5. Schwartz RM, Liu B, Sison C, et al. Study design and results of a population-based study on perceived stress following Hurricane Sandy. *Disaster Med Public Health Prep*. 2016;10(3):325–332.
6. Office of Public Health Preparedness and Response, Centers for Disease Control and Prevention. Public health assessment and surveillance after a disaster. www.cdc.gov/disasters/surveillance/index.html. Published January 3, 2012. Updated September 5, 2014. Accessed August 11, 2016.
7. Brown DL. Disparate effects of the 1989 Loma Prieta and 1994 Northridge earthquakes on hospital admissions for acute myocardial infarction: importance of superimposition of triggers. *Am Heart J*. 1999;137(5):830–836.
8. Swerdel JN, Janevic TM, Cosgrove NM, et al. The effect of Hurricane Sandy on cardiovascular events in New Jersey. *J Am Heart Assoc*. 2014;3(6):e001354.
9. Kim H, Schwartz RM, Hirsch J, et al. Effect of Hurricane Sandy on Long Island emergency departments visits. *Disaster Med Public Health Prep*. 2016;10(3):344–350.
10. Tsai S, Hamby T, Chu A, et al. Development and application of syndromic surveillance for severe weather events following Hurricane Sandy, *Disaster Med Public Health Prep*. 2016; 10(3):463–471.
11. Lee DC, Smith SW, Carr BG, et al. Geographic distribution of disaster-specific emergency department use after Hurricane Sandy in New York City. *Disaster Med Public Health Prep*. 2016;10(3):351–361.
12. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality, US Department of Health and Human Services. Overview of the State Inpatient Databases (SID). https://www.hcup-us.ahrq.gov/sidoverview.jsp. Published May 2013. Accessed August 11, 2016.
13. Olfson M, Wang S, Blanco C. National trends in hospital-treated self-harm events among middle-aged adults. *Gen Hosp Psychiatry*. 2015;37(6):613–619.
14. DeGrauw X, Annest JL, Stevens JA, et al. Unintentional injuries treated in hospital emergency departments among persons aged 65 years and older, United States, 2006–2011. *J Safety Res*. 2016;56:105–109.
15. Sabbatini AK, Kocher KE, Basu A, et al. In-hospital outcomes and costs among patients hospitalized during a return visit to the emergency department. *JAMA*. 2016;315(7):663–671.
16. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002;31(2):422–429.
17. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci*. 2000;1(1):31–49.
18. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017;46(1):348–355.
19. American FactFinder, US Census Bureau. Search. (American Community Survey, 2008–2012. ACS 5-year estimates, Table DP05). https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t. Accessed February 5, 2017.
20. Federal Emergency Management Agency Modeling Task Force. FEMA MOTF Hurricane Sandy impact analysis. https://www.arcgis.com/home/item.html?id=307dd522499d4a44a33d7296a5da5ea0. Created October 27, 2012. Updated July 22, 2015. Accessed August 11, 2016.
21. Federal Emergency Management Agency Modeling Task Force. *Hurricane Sandy Impact Analysis*. Washington, DC: Federal Emergency Management Agency; 2013. https://data.femadata.com/MOTF/SOPs/Hurricane/FEMA%20MOTF-Hurricane%20Sandy%20Products%20README%2004182013.pdf. Accessed August 11, 2016.
22. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality, US Department of Health and Human Services. SID database documentation. https://www.hcup-us.ahrq.gov/db/state/siddbdocumentation.jsp. Accessed August 11, 2016.

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022

23. National Center for Health Statistics, Centers for Disease Control and Prevention. *ICD-9-CM Official Guidelines for Coding and Reporting*. Hyattsville, MD: National Center for Health Statistics; 2011. https://www.cdc.gov/nchs/data/icd/icd9cm_guidelines_2011.pdf. Accessed August 11, 2016.
24. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality, US Department of Health and Human Services. HCUP quality control procedures. https://www.hcup-us.ahrq.gov/db/quality.jsp. Accessed August 11, 2016.
25. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016. http://www.R-project.org. Accessed August 11, 2016.
26. Bivand R, Keitt T, Rowlingson B. rgdal: bindings for the geospatial data abstraction library. https://CRAN.R-project.org/package=rgdal. Published May 12, 2016. Accessed August 11, 2016.
27. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1). (doi: 10.18637/jss.v067.i01).
28. Burstein H. Finite population correction for binomial confidence limits. *J Am Stat Assoc*. 1975;70(349): 67–69.
29. Paul M, Held L. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Stat Med*. 2011;30(10):1118–1136.
30. Salmon M, Schumacher D, Höhle M. Monitoring count time series in R: aberration detection in public health surveillance. *J Stat Softw*. 2016;70(10). (doi: 10.18637/jss.v070.i10).
31. Public Health Information Group, Center for Community Health, New York State Department of Health. *New York State Asthma Surveillance Summary Report. October 2013*. Albany, NY: New York State Department of Health; 2013. https://www.health.ny.gov/statistics/ny_asthma/pdf/2013_asthma_surveillance_summary_report.pdf. Accessed August 11, 2016.
32. Federal Emergency Management Agency. *Damage Assessment Operations Manual: A Guide to Assessing Damage and Impact*. Washington, DC: Federal Emergency Management Agency; 2016. http://www.fema.gov/media-library-data/1459972926996-a31eb90a2741e86699ef34ce2069663a/PDAManualFinal6.pdf. Accessed August 11, 2016.
33. Jacobs PD, Duchovny N, Lipton BJ. Changes in health status and care use after ACA expansions among the insured and uninsured. *Health Aff (Millwood)*. 2016;35(7): 1184–1188.

Downloaded from https://academic.oup.com/aje/article/186/11/1290/3860347 by Centers For Disease Control user on 11 February 2022