Hospitalisations for cardiovascular and respiratory disease among older adults living near unconventional natural gas development: a difference-in-differences analysis

Kevin S Trickey, Zihan Chen, Prachi Sanghavi

Lancet Planet Health 2023; 7: e187–96

Department of Public Health Sciences, Biological Sciences Division, University of Chicago, Chicago, IL, USA (K S Trickey BS, Z Chen MPP, P Sanghavi PhD)

Correspondence to: Dr Prachi Sanghavi, Department of Public Health Sciences, Biological Sciences Division, University of Chicago, Chicago, IL 60637, USA
psanghavi@uchicago.edu

Summary

Background During 2008–15, the Marcellus shale region of the US state of Pennsylvania experienced a boom in unconventional natural gas development (UNGD) or "fracking". However, despite much public debate, little is known about the effects of UNGD on population health in local communities. Among other mechanisms, air pollution from UNGD might affect individuals living nearby through cardiovascular or respiratory disease, and older adults could be particularly susceptible.

Methods To study the health impacts of Pennsylvania's fracking boom, we exploited the ban on UNGD in neighbouring New York state. Using 2002–15 Medicare claims, we conducted difference-in-differences analyses over multiple timepoints to estimate the risk of living near UNGD for hospitalisation with acute myocardial infarction (AMI), chronic obstructive pulmonary disease (COPD) and bronchiectasis, heart failure, ischaemic heart disease, and stroke among older adults (aged ≥65 years).

Findings Pennsylvania ZIP codes that started UNGD in 2008–10 were associated with more hospitalisations for cardiovascular diseases in 2012–15 than would be expected in the absence of UNGD. Specifically, in 2015, we estimated an additional 11·8, 21·6, and 20·4 hospitalisations for AMI, heart failure, and ischaemic heart disease, respectively, per 1000 Medicare beneficiaries. Hospitalisations increased even as UNGD growth slowed. Results were robust in sensitivity analyses.

Interpretation Older adults living near UNGD could be at high risk of poor cardiovascular outcomes. Mitigation policies for existing UNGD might be needed to address current and future health risks. Future consideration of UNGD should prioritise local population health.

Funding University of Chicago and Argonne National Laboratories.

Introduction

The extraction of natural gas from tight shale rock using hydraulic fracturing ("fracking"), or unconventional natural gas development (UNGD), has grown rapidly in the USA during the past two decades.[1] However, public oversight has been limited, as major federal acts that protect air and water exempt the UNGD industry.[2,3] Given the frequent proximity of hydraulic fracturing to residential areas, local population health might be affected by water contamination from fracture fluids[2,4] and air pollution through truck exhaust, drill and pump combustion, and completion venting, which can produce fine particulate matter ($PM_{2·5}$), volatile organic compounds, and nitrogen oxides.[5–7] Yet, despite the many potential pollution pathways and the level of public interest, few studies have systematically evaluated the health risks of living near UNGD sites.

Although most previous research in this area has important limitations, it consistently suggests an association between UNGD and poor health outcomes.[8,9] Researchers have detected elevated levels of airborne hydrocarbons and other pollutants, including BTEX (benzene, toluene, ethylbenzene, and xylene) compounds, methylene chloride, and polycyclic aromatic hydrocarbons, near UNGD.[10–14] A recent analysis found higher all-cause mortality for individuals living near and downwind of UNGD.[15] In epidemiological studies, UNGD has been connected with birth outcomes, including preterm birth,[16,17] low birthweight,[18–21] and congenital heart defects.[22] Studies of paediatric and adult asthma have reported increased risk associated with UNGD.[23–25] Likewise, two analyses that examined acute myocardial infarction (AMI)[26] and heart failure[27] also found positive associations with UNGD. More generally, research investigating adult health outcomes has suggested possible associations between UNGD and increased risk of cardiology procedures,[28] pneumonia events,[29] and genitourinary diagnoses,[30] but most of this work has used broad disease categories.

We focused on the northern Marcellus shale region of Pennsylvania, which experienced a fracking boom between 2008 and 2015. Our study combines improvements to the existing literature in four areas. First, our

Research in context

Evidence before this study

Despite the many potential pollution pathways, few studies have systematically evaluated the health risks of living near unconventional natural gas development (UNGD) or "fracking" sites. Researchers have detected elevated levels of airborne hydrocarbons and other pollutants, including BTEX (benzene, toluene, ethylbenzene, and xylene) compounds, methylene chloride, and polycyclic aromatic hydrocarbons, near UNGD. A recent analysis found individuals who lived downwind of UNGD had significantly higher mortality risk. Other studies have indicated an association between UNGD and acute myocardial infarction (AMI), heart failure, and asthma. However, more often, research investigating adult health outcomes has grouped conditions into broad diagnosis categories, and has suggested possible associations between UNGD and a range of outcomes, including increased risk of pneumonia events and genitourinary disease. Finally, in addition to adult health studies, research has also linked UNGD activity to birth outcomes, including preterm birth, low birth weight, and congenital heart defects.

Added value of this study

We exploited a fracking ban in New York state to assess health outcomes of people aged 65 years and older in neighbouring regions of Pennsylvania that experienced a fracking boom between 2008 and 2015. With one exception, researchers have not been able to take advantage of this setup, in part due to the challenges of obtaining data from both states. Further, in recognition of the growing concern in econometrics about identification and interpretation of two-way fixed effects models, we applied a recently developed approach for difference-in-differences analysis with multiple time periods. Finally, we assessed five specific health outcomes, including AMI, chronic obstructive pulmonary disease and bronchiectasis, heart failure, ischaemic heart disease, and stroke, in parallel analyses. Some of these outcomes have not previously been rigorously studied; others were the only outcomes studied in the particular analysis and thus comparisons across conditions have been challenging due to differences in datasets, comparison groups, and modelling approaches. Our findings indicate that in 2015 alone, there were an additional 11·8, 21·6, and 20·4 hospitalisations for AMI, heart failure, and ischaemic heart disease, respectively, per 1000 Medicare beneficiaries, than would have been expected in the absence of UNGD. These results were robust to multiple model specifications and tests of our assumptions.

Implications of all the available evidence

The scientific evidence to date indicates that UNGD is associated with poor health outcomes across a range of diseases in both infants and older adults. Policies that can help mitigate the current and future health risks of existing UNGD are likely needed, and any future consideration of UNGD should take into account the potential health impacts for the local population.

study design exploited the fracking ban in neighbouring New York state, which was formalised in 2015 by the state's Department of Environmental Conservation,[31] but had existed in practice as a moratorium for many years.[32] Thus, our study design compared ZIP codes in Pennsylvania and New York over a 14-year time period before and after the fracking boom. Second, we obtained hospitalisation claims data of 100% of Medicare fee-for-service beneficiaries from both states during 2002–15, which provided high-quality information on older adults, who might be particularly susceptible to disease caused or exacerbated by pollutants.[33] Third, our statistical methods were responsive to growing concerns about identification and interpretation in two-way fixed effects models and instead applied an econometrics approach more suitable for difference-in-differences analysis with multiple time periods.[34–43] Finally, we assessed five specific cardiovascular and respiratory health outcomes for associations with UNGD in the Marcellus shale region, identified by a priori examination of the current literature on air pollution near UNGD sites.[26,44–53]

Methods

Data sources

We obtained UNGD data from the Pennsylvania Department of Environmental Protection Oil and Gas Reports' SPUD Data Report, including the spud date (the first date of drilling) and the location of every UNGD well in Pennsylvania from 2002 to 2015.[54] Since the construction, drilling, and abandonment phases might be major sources of air pollution,[55] we included all drilled UNGD wells even if they were later plugged or inactive. We joined each well location with the Census Bureau's 2015 Cartographic Boundary File of 5-digit ZIP Code Tabulation Areas (ZCTA).[56]

We used hospitalisation data (MedPAR) of 100% of Medicare fee-for-service beneficiaries from Jan 1, 2002, to Dec 31, 2015, and extracted admission dates and up to 25 discharge diagnoses.[57] Medicare is a US national public insurance programme for people aged 65 years or older; younger individuals with specific disabilities and diseases, such as end-stage renal disease, can also qualify, but we did not include these individuals. We obtained each beneficiary's enrolment and demographic information, as well as their mailing ZIP code, from the Master Beneficiary Summary Files. We required beneficiaries to be enrolled in Medicare Parts A and B during the month of their admission. Additional county-level socioeconomic and health utilisation indicators were obtained from the Small Area Income and Poverty Estimates programme and intercensal population estimates for 2002 and 2015.

Exposed and unexposed regions

We selected 36 ZIP codes in Bradford, Susquehanna, and Tioga counties in northern Pennsylvania as exposed areas based on the beginning of UNGD between 2008 and 2010 (figure 1, appendix p 6). Specifically, we included 16 ZIP codes that started UNGD in 2008 and ten each in 2009 and 2010. We identified two unexposed neighbouring regions in New York state with similar demographics and rural economies as the Pennsylvania region. These New York regions experienced no UNGD between 2002 and 2015, due to a statewide moratorium, and later ban, on fracking permits. To minimise "off-target" UNGD exposure (eg, air pollution or truck traffic from Pennsylvania) in the control ZIP codes, we selected as our main control the New York region (60 ZIP codes) that does not directly border Pennsylvania (the northern of the two regions in figure 1). The appendix (pp 2, 10) includes analysis with the second New York region (68 ZIP codes) that directly borders the exposed area.

Health outcome measures

We selected five specific health outcomes based on strength of evidence in the academic literature for possible short-term air pollution effects. Existing literature on $PM_{2·5}$, volatile organic compounds, and general traffic-related pollution[44] suggested chronic obstructive pulmonary disease (COPD)[45–48] and bronchiectasis,[58–60] general ischaemic heart disease,[6,45,49,50] acute myocardial infarction (AMI),[26,48] stroke,[45,51–53] and heart failure[6,27,45,48,50] as likely candidates for UNGD-related health outcomes. We applied Chronic Conditions Data Warehouse (CCW) algorithms to identify these outcomes through International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis codes (appendix pp 2, 7–8).[57,61] Since the CCW algorithms grouped COPD and bronchiectasis together, we also joined these two diagnoses under the category of COPD. The switch from ICD-9-CM to ICD-10-CM occurred on Oct 1, 2015, affecting the last 3 months of our dataset. While AMI cases were a particular subset of ischaemic heart disease cases, the other outcomes were defined by mutually exclusive codes.

We separately analysed cases that had a primary diagnosis, defined by a diagnosis in the first two diagnosis code fields, from cases in which any of the 25 diagnosis codes identified the condition. Primary diagnoses indicate a leading reason for the hospitalisation, whereas secondary diagnoses might capture comorbid conditions or even past occurrences. We tabulated the number of hospitalisations for each outcome by ZIP code and year, as determined by the admission date.

Statistical analysis

We applied difference-in-differences (DID) analysis for multiple timepoints as described by Callaway and colleagues.[34] This approach avoids the identification and

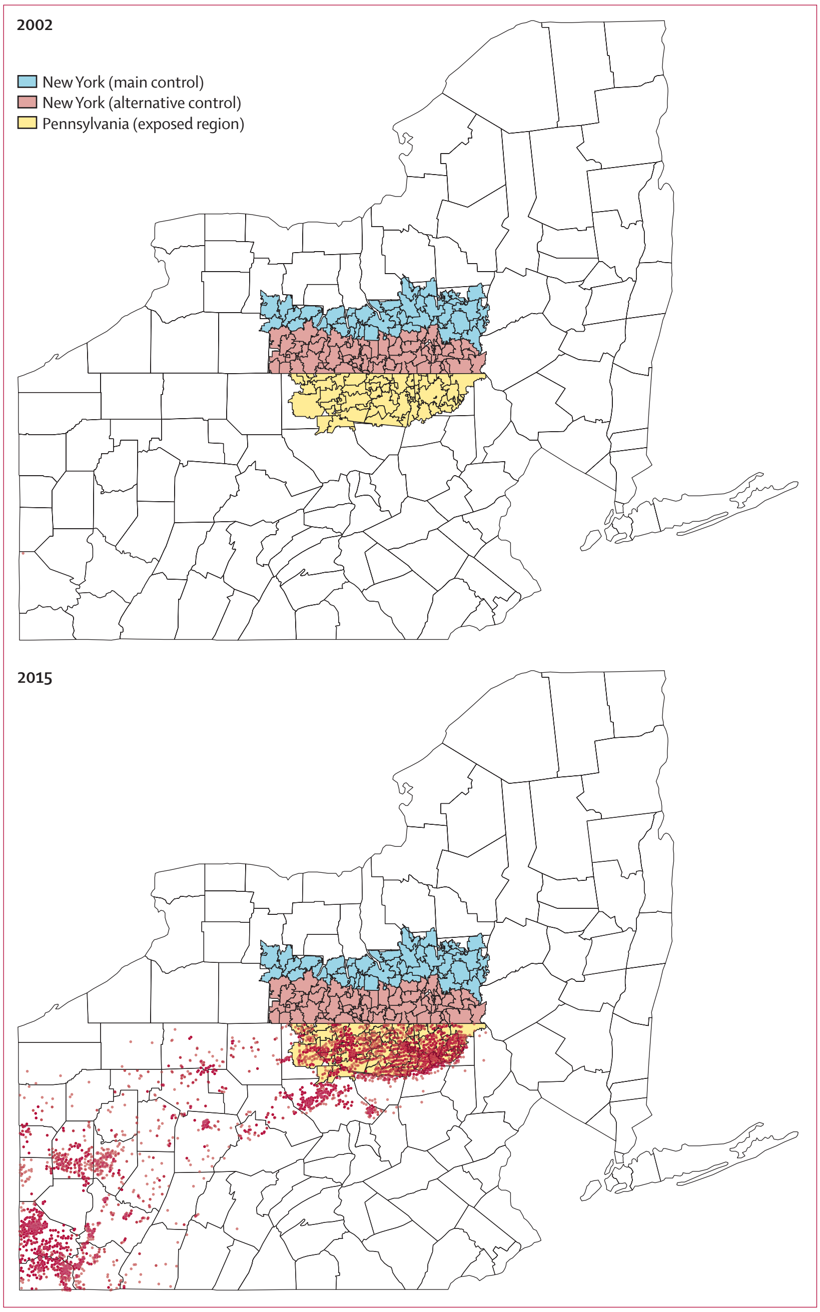

Figure 1: **UNGD wells by region, 2002 and 2015**
Maps of Pennsylvania and New York show the locations of drilled UNGD wells (red dots) in Pennsylvania by the end of 2002 (one total well) and by Dec 31, 2015, the end of our study period (9596 total wells). The southernmost group of ZIP codes in Pennsylvania (yellow) were all exposed to UNGD activity and the 36 of these that started UNGD between 2008 and 2010 were included in the study. The two northern groups, labelled the main (blue, further north) and alternative (red) control regions, were UNGD-free, due to a statewide ban on UNGD, although the alternate control might have experienced air pollution-related exposure due to proximity. ZIP codes that were included in the study are listed in the appendix (p 6). UNGD=unconventional natural gas development.

interpretation challenges of the commonly used two-way fixed effects models that are now being recognised in the econometrics literature.[35,36,38–43] Fundamentally, these papers argue that because two-way fixed effects models average across comparisons in multiple directions

See Online for appendix

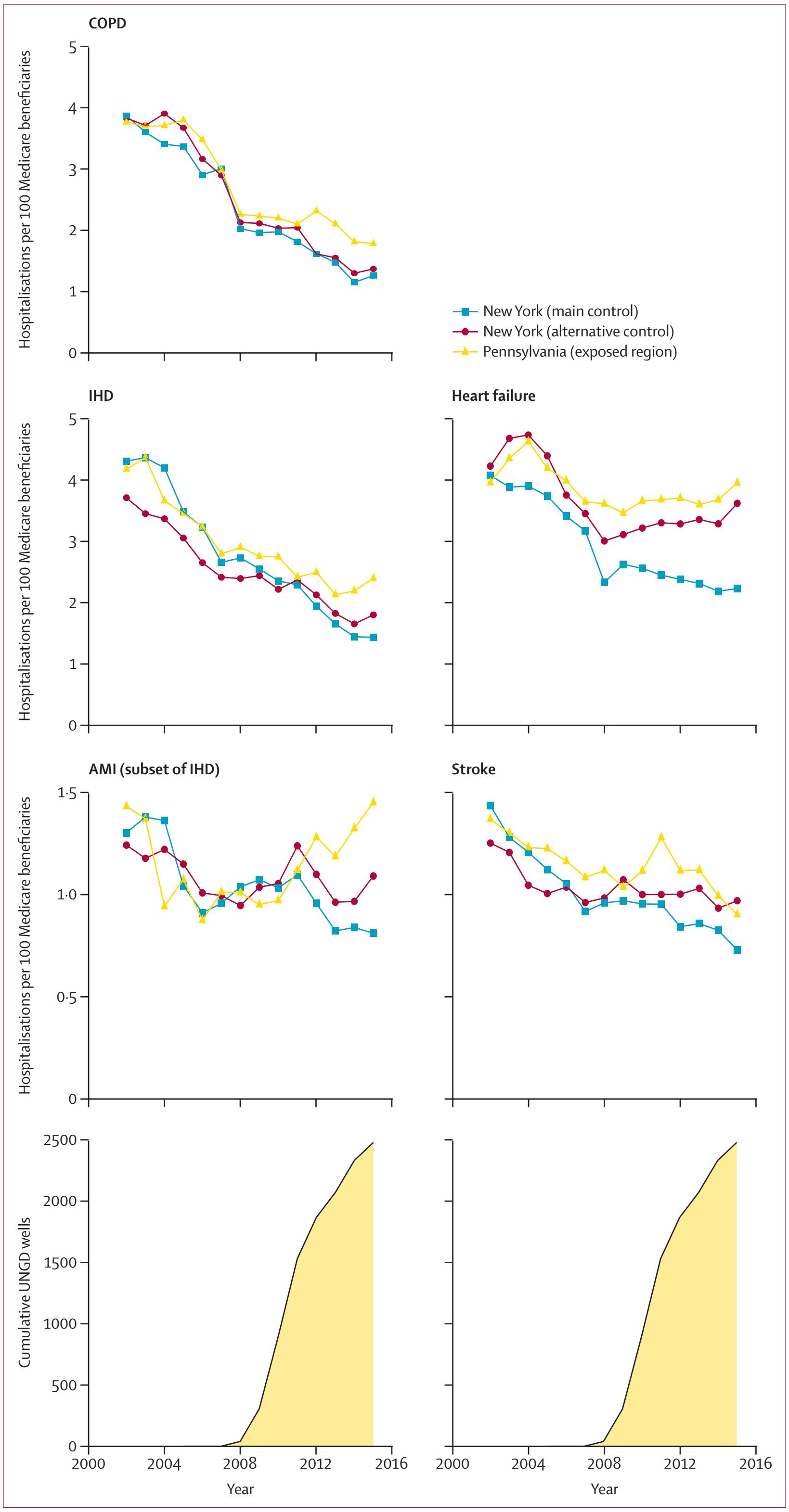

Figure 2: **Hospitalisation rates by disease and UNGD activity by study region, 2002–15**

Hospitalisations for each outcome per 100 beneficiaries across all ZIP codes in each study region are plotted for each year, 2002–15. Only cases identified using primary diagnosis codes (first two codes) are included. In 2015, we used both ICD-9-CM and ICD-10-CM codes to identify each health outcome. Yellow triangles indicate hospitalisation rates among the 36 Pennsylvania ZIP codes included in the study, which were exposed to UNGD (shown in the bottom plots). Blue squares and red circles represent hospitalisation rates in the main New York and alternative New York control regions, respectively. The alternative control region directly borders the exposed ZIP codes and might be partially affected by Pennsylvania UNGD activity. AMI cases were a subset of IHD cases. AMI=acute myocardial infarction. COPD=chronic obstructive pulmonary disease. IHD=ischaemic heart disease. UNGD=unconventional natural gas development.

(within-unit across time, across-unit within time, across-unit across time), conceiving the counterfactual scenarios under study is difficult. The methods proposed by Callaway and colleagues extend the usual DID analysis, which can have a clear interpretation when appropriately used, and are increasingly being used in economics research.[37,62–66]

Specifically, for each outcome, we conducted multiple DID analyses across which the treatment and control groups, as well as the pre-period, stayed fixed, but the post-period shifted forward 1 year at a time. Thus, each linear regression model took the standard form:

$$\text{hospitalisation}_{zt} = \alpha\ \text{treated}_z + \gamma\ \text{post}_t + \beta(\text{treated}_z \times \text{post}_t) + \varepsilon_{zt}$$

where hospitalisation$_{zt}$ was the per 100 beneficiary count of hospitalisations in ZIP code z and year t for the diagnosis, treated$_z$ was a binary indicator of UNGD activity in the ZIP code, post$_t$ was an indicator for whether the year was in the pre-period or post-period, and ε_{zt} was the error term. The model was weighted by the Medicare population of each ZIP code. Our quantity of interest was β, interpreted as the change in hospitalisation rate between the pre-period and post-period in Pennsylvania beyond what would be expected in the absence of UNGD. Although we used linear models for our main analyses to allow easier interpretation, the appendix (pp 2, 11) includes negative binomial models, which are often used in air pollution studies,[67] with hospitalisations specified as counts. Additionally, the appendix (pp 3, 12) includes results adjusted for multiple testing using a multiplier bootstrap procedure.

Although exposure was defined based on starting UNGD in 2008, 2009, or 2010, relatively few wells were actually developed in 2008 and 2009 (see figure 2). Since exposure would have been minimal in those years, we chose to consider 2009 as the pre-period point for all 36 ZIP codes in our main analyses. One large, exposed group instead of three smaller groups starting treatment at different times allowed us to use the data more efficiently and only made our estimates more conservative should there have been some exposure in the pre-period.

	2002			2015		
	Pennsylvania (exposed region)	New York (main control)	New York (alternative control)	Pennsylvania (exposed region)	New York (main control)	New York (alternative control)
UNGD activity						
Annual new wells	0	0	0	146	0	0
Cumulative wells	0	0	0	2473	0	0
Total land area, square miles	2194	2601	2790	2194	2601	2790
Characteristics of study population: Medicare beneficiaries						
Number of beneficiaries	16 703	31 978	72 468	21 119	40 033	82 328
Percent male	44·9%	43·2%	42·4%	46·7%	45·2%	44·6%
Mean age, years (SD)	72·8 (11·6)	72·9 (12·0)	73·1 (11·9)	71·9 (11·8)	72·0 (12·1)	71·9 (12·8)
Percent White	99·1%	97·8%	97·4%	97·9%	95·2%	94·8%
Percent dual-eligible with Medicaid*	20·3%	18·8%	19·6%	20·0%	22·2%	23·7%
County-level characteristics						
Counties used	Bradford, Susquehanna, Tioga	Chenango, Cortland, Schuyler, Tompkins	Broome, Chemung, Steuben, Tioga	Bradford, Susquehanna, Tioga	Chenango, Cortland, Schuyler, Tompkins	Broome, Chemung, Steuben, Tioga
Total population	146 633	217 502	443 626	144 951	218 099	429 396
Percent older than 65 years	15·8%	12·0%	15·5%	20·5%	15·4%	17·7%
Poverty rate	11·2%	11·9%	11·9%	12·9%	16·0%	15·9%
Median income (2002 dollars)†	$34 457	$35 920	$36 413	$37 171	$39 621	$36 801

Characteristics are presented comparing the different study regions defined in figure 1. UNGD data were obtained from the Pennsylvania Department of Environmental Conservation. Demographic characteristics of the study population were extracted from Medicare claims data, which are also used for outcome measurements. County-level characteristics were obtained from the Small Area Income and Poverty Estimates (SAIPE) programme and intercensal population estimates. County and ZIP code geographies are not exact overlaps, so the county-level characteristics reported here are approximations to our study regions only. UNGD=unconventional natural gas development. *The percent of beneficiaries dual-eligible with Medicaid was not available to us before 2006, so values reported under 2002 are for 2006. †Median incomes are reported as the population-weighted median income across the three or four counties.

Table 1: **UNGD activity, Medicare population, and county-level characteristics of exposed and unexposed study regions**

Nonetheless, the appendix (pp 3, 13–18) includes analysis of the three groups using the group-time average treatment effects methods by Callaway and colleagues.

For our context, in which UNGD and thus potential pollution exposure increased each year, this methodological approach was especially fitting because it allowed assessment of whether any observed increases in hospitalisation rates were sustained or increased in the face of UNGD growth. If so, it might build confidence in an association between hospitalisation and UNGD, while a decreasing rate would diminish it.

Finally, the DID analysis rests on the assumption that trends in hospitalisation rates would be parallel between the exposed and unexposed ZIP codes in the absence of UNGD. To test this assumption, we conducted two sets of DID analyses.[68] In the main parallel trends analysis, we assessed changes between Pennsylvania and New York hospitalisation rates for each condition between 2009 (the post-period) and each year in turn between 2002 and 2008 (the pre-period). In the second analysis, included in the appendix (pp 4, 19), we compared each year in the pre-period with the previous year—eg, 2003 versus 2002, 2004 versus 2003.

Data analysis was performed in Python 3.7.6, R 3.5.0, and R 4.1.2 for the *did* package, version 2.1.1. All code and non-identifiable data are available online and are described in the appendix (pp 20–22).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Northern Pennsylvania experienced rapid UNGD growth during the study period (figure 1). The first UNGD in these ZIP codes was in 2005, with only nine total spuds by 2007. However, 69 new spuds were developed in 2008, and between 217 and 861 spuds were drilled annually for the next 7 years, with peak drilling in 2010 (764) and 2011 (861).

Although DID analysis requires parallel trends between treatment and control groups in the absence of treatment, it does not explicitly require groups to be similar otherwise. However, the parallel trends assumption is much more plausible if the groups share similar levels and trends of observed characteristics. Table 1 shows average characteristics of the treatment and control ZIP codes for 2002 and 2015, the beginning and end of our study period. By 2015, there were on average 1·13 UNGD

	Primary diagnoses*					Any diagnosis†				
	AMI	COPD and bronchiectasis	Heart failure	Ischaemic heart disease (including AMI)	Stroke	AMI	COPD and bronchiectasis	Heart failure	Ischaemic heart disease (including AMI)	Stroke
2010	0·06 (–0·22 to 0·34)	–0·05 (–0·45 to 0·35)	0·26 (–0·24 to 0·77)	0·18 (–0·32 to 0·69)	0·09 (–0·18 to 0·36)	0·03 (–0·26 to 0·32)	0·23 (–0·39 to 0·85)	0·37 (–0·47 to 1·21)	0·60 (–0·24 to 1·45)	0·11 (–0·19 to 0·41)
2011	0·14 (–0·22 to 0·51)	0·03 (–0·45 to 0·51)	0·41 (–0·22 to 1·04)	–0·09 (–0·59 to 0·42)	0·26 (–0·05 to 0·56)	0·19 (–0·17 to 0·55)	0·78 (–0·10 to 1·67)	0·73 (–0·28 to 1·74)	1·31‡ (0·24 to 2·38)	0·26 (–0·02 to 0·54)
2012	0·44‡ (0·12 to 0·77)	0·44 (–0·02 to 0·90)	0·48 (–0·33 to 1·30)	0·35 (–0·34 to 1·03)	0·20 (–0·05 to 0·46)	0·53‡ (0·17 to 0·88)	1·23‡ (0·36 to 2·10)	1·37‡ (0·23 to 2·52)	1·90‡ (0·56 to 3·25)	0·23 (–0·08 to 0·54)
2013	0·49‡ (0·23 to 0·75)	0·37 (–0·17 to 0·90)	0·45 (–0·30 to 1·20)	0·27 (–0·30 to 0·85)	0·20 (–0·08 to 0·48)	0·80‡ (0·49 to 1·12)	1·89‡ (0·79 to 2·99)	1·91‡ (0·73 to 3·09)	2·20‡ (0·82 to 3·58)	0·18 (–0·10 to 0·47)
2014	0·61‡ (0·29 to 0·93)	0·38 (–0·11 to 0·87)	0·63 (–0·22 to 1·49)	0·55 (–0·01 to 1·11)	0·10 (–0·17 to 0·38)	0·91‡ (0·52 to 1·30)	1·24‡ (0·13 to 2·34)	1·74‡ (0·57 to 2·91)	1·19‡ (0·15 to 2·22)	0·16 (–0·14 to 0·46)
2015§	0·77‡¶ (0·45 to 1·10)	0·23 (–0·26 to 0·72)	0·88‡ (0·16 to 1·60)	0·77‡ (0·17 to 1·37)	0·11 (–0·20 to 0·42)	1·18‡ (0·81 to 1·56)	0·78 (–0·25 to 1·81)	2·16‡ (1·10 to 3·22)	2·04‡ (1·14 to 2·94)	0·12 (–0·19 to 0·43)

95% CIs are shown in brackets. Each health outcome was measured by the number of hospitalisations divided by the total Medicare fee-for-service population within the ZIP code and multiplied by 100. The control region was the main New York control in this table. The model was weighted by the Medicare population of each ZIP code. AMI=acute myocardial infarction. COPD=chronic obstructive pulmonary disease. UNGD=unconventional natural gas development. *Health outcomes were determined by the first two diagnosis codes in MedPAR hospital admissions data. †Health outcomes were determined by all diagnosis codes in MedPAR hospital admissions data. ‡Estimates are statistically significant at an alpha level of 0·05. §The outcome values in 2015 were determined by both ICD-9-CM and ICD-10-CM codes due to the transition to ICD-10-CM on Oct 1, 2015. ¶Estimate can be interpreted as: in 2015, compared with 2009, exposed ZIP codes in Pennsylvania experienced 7·7 additional hospitalisations for AMI per 1000 Medicare beneficiaries than would have been expected in the absence of UNGD.

Table 2: **Difference-in-differences analysis: associations between hospitalisation rates and UNGD in Pennsylvania, in each post-treatment year compared with 2009, relative to changes over the same time period in New York**

wells per square mile in the exposed region. Over this time, the population of Medicare beneficiaries in the exposed region grew by 26% and in the main control region by 25%. Both regions experienced similar, slight demographic shifts including increases in male representation of about 2 percentage points, decreases in the White majority by 1–3 percentage points, and decreases in average beneficiary age by about 1 year. County-level indicators also show similar trends in all-age population (0–3% decrease), percentage of population older than 65 years (2–5 percentage point increase), and poverty rate (2–4 percentage point increase).

Hospitalisation rates for our diagnosis groups followed relatively parallel decreasing trends between 2002 and 2008, the beginning of the UNGD boom in the Pennsylvania region (figure 2). After 2009, the hospitalisation rates of the three regions appear to diverge, with the highest hospitalisation rates in the exposed region and the lowest in the main control region in all diagnosis groups. However, formal analysis is required to know if these differences are statistically significant.

Table 2 shows the results of the DID analyses, separately for cases identified using primary diagnosis codes only and any diagnosis code. For AMI, the association between hospitalisation rates and UNGD grew consistently in magnitude between 2012 and 2015 in both primary and any diagnosis cases. Specifically, among primary diagnosis cases, we estimated an additional 4·4 and 7·7 hospitalisations per 1000 Medicare beneficiaries in 2012 and 2015 (relative to 2009), respectively, than would be expected in the absence of UNGD; among any diagnosis cases, these estimates are 5·3 and 11·8 hospitalisations. We do not find any statistically significant associations between hospitalisation rates and UNGD for COPD determined by primary diagnosis. Although COPD hospitalisation rates using any diagnosis code do seem to exhibit an increase over the expected rates between 2012 and 2014, these estimates waver in magnitude and by 2015 are not statistically significant.

In the cases of both heart failure and ischaemic heart disease, 2015 hospitalisation rates indicate a positive association with UNGD when only primary diagnoses are considered. Between 2013 and 2015, the magnitudes of the associations increase with each year, but are not statistically significant. However, when any diagnosis code is included, which substantially increases sample size (appendix pp 2, 9), hospitalisation rates in each year between 2012 and 2015 are higher by between 13·7 to 21·6 cases per 1000 beneficiaries for heart failure and 11·9 to 22·0 cases for ischaemic heart disease than would be expected in the absence of UNGD, albeit with some fluctuation between years. We did not observe any indication of an association between stroke-related hospitalisations and UNGD.

The indications of an association between hospitalisation rates and UNGD in the cases of AMI, heart failure, and ischaemic heart disease warrant further examination in the pre-period. Figure 3 shows the results of repeating the DID approach from the main analyses, but with 2009 as the post-period and each year between 2002 and 2008 as the pre-period in turn. The AMI analysis in 2002 (any diagnosis) and 2004 (any diagnosis and primary diagnosis), the heart failure analysis in 2002 (primary diagnosis), and the ischaemic heart disease analysis in 2005 (any diagnosis) and 2004 (primary diagnosis) all have statistically significant associations. The appendix (p 19) contains an additional parallel trends analysis suggested by Callaway and colleagues,[34] in which pairs of years were assessed, and shows only statistically significant associations in the 2004–05 AMI and the 2003 and 2008 heart failure analyses. Altogether, these findings do not threaten the validity of our main results.

The appendix contains additional sensitivity analyses. The comparison with the New York alternative control (appendix p 10), which we were concerned might also have been exposed to UNGD-related air pollution, generally shared the same direction of results as with the New York main control. However, with the exception of AMI, these hospitalisation rates were not statistically different from Pennsylvania's rates between the pre-period and post-period. Our findings for AMI and heart failure when applying a negative binomial regression model (appendix p 11) were similar to the main results, while our estimates for ischaemic heart disease were more sensitive. Adjustment for multiple testing (appendix p 12) and estimating group-time average treatment effects (appendix pp 13–18) produced results similar to our main findings.

Discussion

UNGD had a strong association, robust to multiple sensitivity analyses, with increased hospitalisation for AMI among Medicare beneficiaries in the Marcellus shale region of Pennsylvania, relative to neighbouring regions in New York where fracking was banned. Specifically, in 2015, there were 11·8 more AMI-related hospitalisations per 1000 beneficiaries in areas with UNGD compared with 2009, than would be expected in the absence of the fracking boom during this time. We also found strong indications of associations between UNGD and heart failure and ischaemic heart disease, although these results exhibited some sensitivity to alternative specifications and assumptions. UNGD was not associated with COPD or stroke. Our study indicates that impacts on local population health should be high priority in the management of existing wells and a key factor in consideration of future UNGD.

Our study design had important strengths. We exploited a unique policy of New York state, a fracking ban, to study

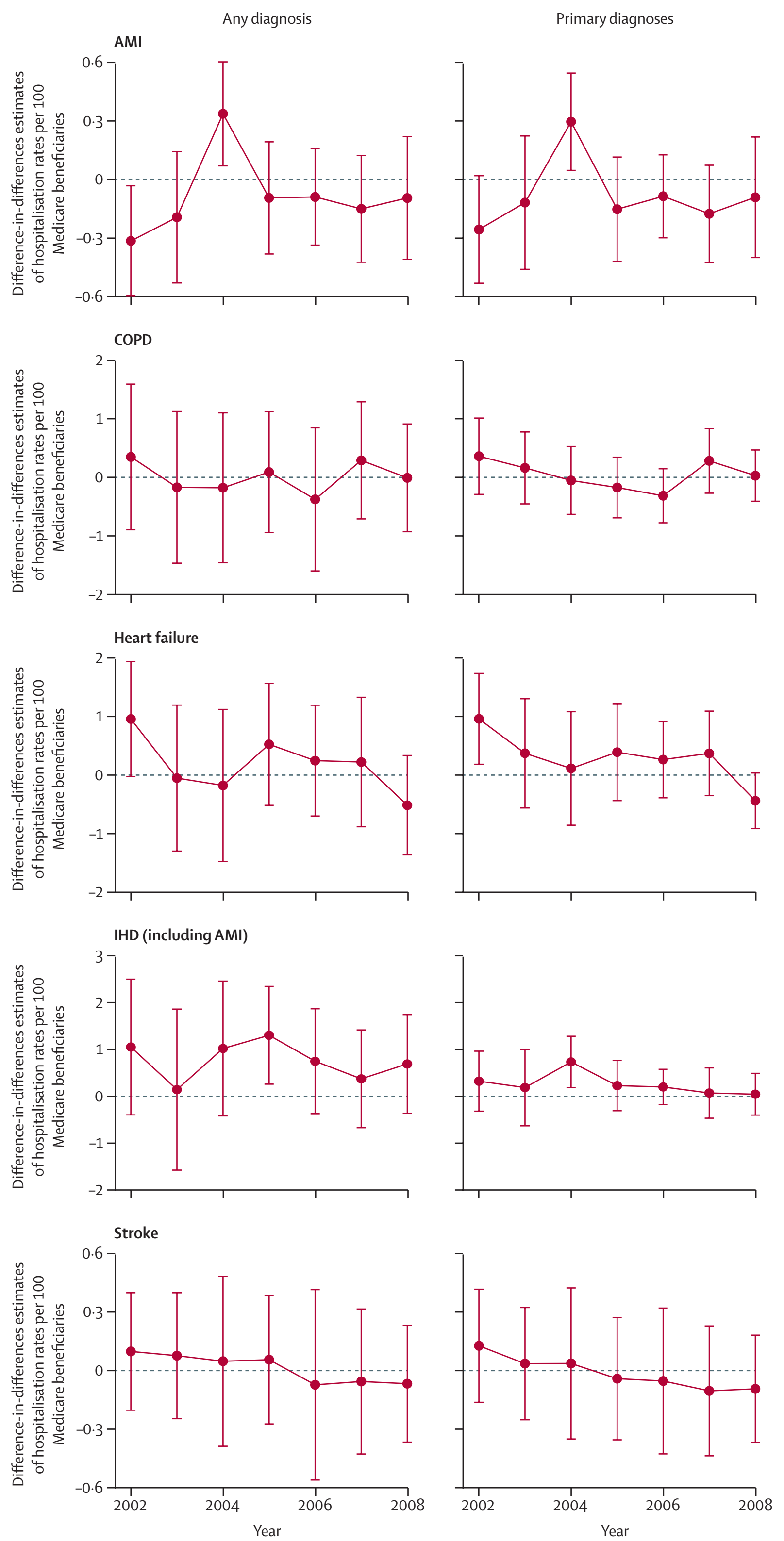

Figure 3: **Pre-period parallel trends analysis, using difference-in-differences analysis of hospitalisation rates in Pennsylvania *vs* New York (main control), between 2009 (the post-period) and each year between 2002 and 2008 (the pre-period)**

Confidence intervals that include 0 indicate no statistically significant difference in hospitalisation rates could be estimated by these two sets of simultaneous comparisons (Pennsylvania *vs* New York and 2009 *vs* the pre-period year). AMI=acute myocardial infarction. COPD=chronic obstructive pulmonary disease. IHD=ischaemic heart disease.

outcomes in neighbouring Pennsylvania. The 100% sample of Medicare fee-for-service beneficiaries gave us high-quality, consistent information about outcomes from both sides of the state line. The long timespan of our analysis, which included the explosive period of UNGD growth, allowed an in-depth assessment of the parallel trends assumption in our methodological approach. Finally, we applied DID analysis to multiple timepoints using methods that avoid the identification and interpretation problems of two-way fixed effects models that are gaining attention in the econometrics literature.[34–42] This DID approach fit the problem at hand, which included increasing levels of exposure to UNGD over time. It also allowed us to base our inference on analysis of both sampling uncertainty and patterns of association over time. Hence, we did not draw strong conclusions about COPD, even though multiple associations between COPD hospitalisations and UNGD were statistically significant.

Our findings have similarities and differences with those from the relatively few previous investigations of UNGD and chronic disease outcomes. A study by Peng and colleagues reported significant increases in pneumonia and some model-sensitive increases in AMI and COPD associated with UNGD, among other outcomes.[29] However, Peng and colleagues used as controls all non-exposed Pennsylvania counties, including urban and suburban centres like Philadelphia and Harrisburg, which might experience differential trends over time in characteristics compared with rural areas. In contrast, our control areas were demographically and economically similar to the largely rural exposed areas. Our departure from the traditional econometric methods for panel data analysis, used by Peng and colleagues, could also partly explain our different findings.

A more recent analysis assessed just AMI outcomes using state hospital discharge data from Pennsylvania and New York and offers a comparison point for our AMI results based on Medicare claims.[26] In their work, Denham and colleagues also find statistically significant and positive associations between AMI hospitalisation rates and UNGD, but methodological differences make precise comparison of magnitudes difficult. Denham and colleagues applied a two-way fixed effects model we previously discussed. There were also differences in their treatment specification, the identification of treatment and control groups, and examinations of time. In an epidemiological study, researchers applied case-control methods to investigate the association between heart failure and UNGD, using electronic health records data from a Pennsylvania health system.[27] Here too, despite differences in datasets, populations, and methodological approaches that make precise comparisons challenging, both of our analyses indicate people living near UNGD are at higher risk of heart failure hospitalisation.

Our estimates for AMI hospitalisation rates steadily increased from 2012 to 2015, suggesting that the increasing number of wells over time might be associated with wider divergence of AMI hospitalisations in Pennsylvania compared with unexposed areas. More generally, AMI, heart failure, and ischaemic heart disease hospitalisation rates in Pennsylvania continued to diverge from those of the unexposed regions even in years past peak development. The initial construction, drilling, and fracturing phases of UNGD are thought to generate the most air pollution and generally occur well within 1 year of the spud date.[5,69] However, these findings would suggest pollutants from operational wells, which involve trucking for production and waste disposal, also play a significant role in poor health outcomes. Therefore, existing wells should not be ignored in UNGD-related policy development, especially as the number of abandoned, sometimes unplugged wells increases amid UNGD bankruptcies.[70]

This study has limitations. We only analysed inpatient admissions, which generally represent only the most serious cases, as opposed to outpatient or physician visit claims. Likewise, our study assessed air pollution-related health effects in the short term, without consideration of potential long-term ramifications of UNGD-related air pollution on these same outcomes, other diseases such as cancer, or conditions more closely related with water contamination.[71] Likewise, we did not quantify magnitudes of exposure, although our assessment of changes in association between health outcomes and UNGD over time offers some limited insight that we previously described. We also did not study specific mechanisms or interventions. Thus, although we focus on air pollution as a mechanism that would be consistent with the outcomes that we studied, it is possible that the health associations we observed were produced in part by other downstream consequences of UNGD.[72]

Although we removed time-invariant and ZIP code-invariant confounders, it is possible that a time-varying ZIP code characteristic could have confounded our results. However, we are not aware of a characteristic that would produce the patterns we observed with AMI, without being a consequence of UNGD itself (such as income or employment rates)[72] and thus a post-treatment effect. While some have suggested that ZIP code-level measures might be subject to confounding from demographic changes associated with a migrating UNGD workforce, it is unlikely that adults older than 65 years would be part of this migration.[29,30] We also investigated New York state policies that might have led to concurrent changes in hospitalisation rates and found one possibility, a trans-fatty acids restriction in eateries.[73] However, this policy was not implemented in any counties in our main control and only one county in the alternative control area. Finally, since well locations were based on ZCTAs and hospitalisations based on ZIP codes, some misalignment could have occurred. However, we used a binary specification for treatment status and any such misalignment would only have occurred within the treatment area, in which most ZIP codes had at least one well.

Despite the level of public debate and concern, few robust academic studies have assessed the associations between local population health and UNGD. We found Medicare beneficiaries in northern Pennsylvania experienced negative cardiovascular health outcomes associated with rapid and dense UNGD in the region. Importantly, hospitalisation rates grew even as UNGD was slowing, suggesting that operational and abandoned wells might be significant sources of pollution. Policy makers should urgently address the safety of existing wells, including by surveilling local health, and prioritise local population health when considering any new UNGD.

Contributors

KST contributed to data curation, formal analysis, methodology, coding, validation, visualisation, and writing the original draft. ZC contributed to formal analysis, methodology, coding, validation, visualisation, and reviewing and editing the manuscript. PS contributed to conceptualisation, data curation, formal analysis, funding acquisition, methodology, project administration, supervision, validation, visualisation, and reviewing and editing the manuscript. All authors had full access to all the data in the study, verified the data, and had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

All code and some de-identified data are available at https://github.com/sanghavi-lab/medicare_claims_and_pennsylvania_UNGD-. Further data may be shared upon request to the corresponding author, subject to the cell size suppression policy of the US Centers for Medicare and Medicaid Services.

Acknowledgments

This study was funded by a grant from the University of Chicago and Argonne National Laboratories. The funders had no role in the writing of the manuscript or the decision to submit it for publication.

References

1 U.S. Energy Information Administration. Hydraulically fractured horizontal wells account for most new oil and natural gas wells. Jan 30, 2018. https://www.eia.gov/todayinenergy/detail.php?id=34732 (accessed Oct 28, 2020).

2 Trickey K, Hadjimichael N, Sanghavi P. Public reporting of hydraulic fracturing chemicals in the USA, 2011–18: a before and after comparison of reporting formats. *Lancet Planet Health* 2020; **4:** e178–85.

3 Gorski I, Schwartz BS. Environmental health concerns from unconventional natural gas development. In: Oxford Research Encyclopedia of Global Public Health. Oxford University Press, 2019. https://doi.org/10.1093/acrefore/9780190632366.013.44 (accessed Feb 12, 2023).

4 Landrigan PJ, Frumkin H, Lundberg BE. The false promise of natural gas. *N Engl J Med* 2020; **382:** 104–07.

5 Roy AA, Adams PJ, Robinson AL. Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. *J Air Waste Manag Assoc* 2014; **64:** 19–37.

6 Ye D, Klein M, Chang HH, et al. Estimating acute cardiorespiratory effects of ambient volatile organic compounds. *Epidemiology* 2017; **28:** 197–206.

7 Martinelli N, Olivieri O, Girelli D. Air particulate matter and cardiovascular disease: a narrative review. *Eur J Intern Med* 2013; **24:** 295–302.

8 Deziel NC, Brokovich E, Grotto I, et al. Unconventional oil and gas development and health outcomes: a scoping review of the epidemiological research. *Environ Res* 2020; **182:** 109124.

9 Werner AK, Vink S, Watt K, Jagals P. Environmental health impacts of unconventional natural gas development: a review of the current strength of evidence. *Sci Total Environ* 2015; **505:** 1127–41.

10 Colborn T, Schultz K, Herrick L, Kwiatkowski C. An exploratory study of air quality near natural gas operations. *Hum Ecol Risk Assess* 2014; **20:** 86–105.

11 Macey GP, Breech R, Chernaik M, et al. Air concentrations of volatile compounds near oil and gas production: a community-based exploratory study. *Environ Health* 2014; **13:** 82.

12 Pétron G, Frost G, Miller BR, et al. Hydrocarbon emissions characterization in the Colorado Front Range: a pilot study. *J Geophys Res* 2012; **117:** D04304.

13 Rich AL, Orimoloye HT. Elevated atmospheric levels of benzene and benzene-related compounds from unconventional shale extraction and processing: human health concern for residential communities. *Environ Health Insights* 2016; **10:** 75–82.

14 Vinciguerra T, Yao S, Dadzie J, et al. Regional air quality impacts of hydraulic fracturing and shale natural gas activity: evidence from ambient VOC observations. *Atmos Environ* 2015; **110:** 144–50.

15 Li L, Dominici F, Blomberg AJ, et al. Exposure to unconventional oil and gas development and all-cause mortality in Medicare beneficiaries. *Nat Energy* 2022; **7:** 177–85.

16 Whitworth KW, Marshall AK, Symanski E. Maternal residential proximity to unconventional gas development and perinatal outcomes among a diverse urban population in Texas. *PLoS One* 2017; **12:** e0180966.

17 Casey JA, Savitz DA, Rasmussen SG, et al. Unconventional natural gas development and birth outcomes in Pennsylvania, USA. *Epidemiology* 2016; **27:** 163–72.

18 Currie J, Greenstone M, Meckel K. Hydraulic fracturing and infant health: new evidence from Pennsylvania. *Sci Adv* 2017; **3:** e1603021.

19 Stacy SL, Brink LL, Larkin JC, et al. Perinatal outcomes and unconventional natural gas operations in Southwest Pennsylvania. *PLoS One* 2015; **10:** e0126425.

20 Hill EL. Shale gas development and infant health: evidence from Pennsylvania. *J Health Econ* 2018; **61:** 134–50.

21 Willis MD, Hill EL, Boslett A, Kile ML, Carozza SE, Hystad P. Associations between residential proximity to oil and gas drilling and term birth weight and small-for-gestational-age infants in Texas: a difference-in-differences analysis. *Environ Health Perspect* 2021; **129:** 77002.

22 McKenzie LM, Guo R, Witter RZ, Savitz DA, Newman LS, Adgate JL. Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. *Environ Health Perspect* 2014; **122:** 412–17.

23 Willis M, Hystad P, Denham A, Hill E. Natural gas development, flaring practices and paediatric asthma hospitalizations in Texas. *Int J Epidemiol* 2021; **49:** 1883–96.

24 Rasmussen SG, Ogburn EL, McCormack M, et al. Association between unconventional natural gas development in the marcellus shale and asthma exacerbations. *JAMA Intern Med* 2016; **176:** 1334–43.

25 Willis MD, Jusko TA, Halterman JS, Hill EL. Unconventional natural gas development and pediatric asthma hospitalizations in Pennsylvania. *Environ Res* 2018; **166:** 402–08.

26 Denham A, Willis MD, Croft DP, Liu L, Hill EL. Acute myocardial infarction associated with unconventional natural gas development: a natural experiment. *Environ Res* 2021; **195:** 110872.

27 McAlexander TP, Bandeen-Roche K, Buckley JP, et al. Unconventional natural gas development and hospitalization for heart failure in Pennsylvania. *J Am Coll Cardiol* 2020; **76:** 2862–74.

28 Jemielita T, Gerton GL, Neidell M, et al. Unconventional gas and oil drilling is associated with increased hospital utilization rates. *PLoS One* 2015; **10:** e0131093.

29 Peng L, Meyerhoefer C, Chou SY. The health implications of unconventional natural gas development in Pennsylvania. *Health Econ* 2018; **27:** 956–83.

30 Denham A, Willis M, Zavez A, Hill E. Unconventional natural gas development and hospitalizations: evidence from Pennsylvania, United States, 2003–2014. *Public Health* 2019; **168:** 17–25.

31 New York State Department of Environmental Conservation. Final Supplemental Generic Environmental Impact Statement on the Oil, Gas, and Solution Mining Regulatory Program. June, 2015. https://www.dec.ny.gov/docs/materials_minerals_pdf/findingstatehvhf62015.pdf (accessed Oct 28, 2020).

32 Crowell & Moring LLP. New York issues statewide fracking ban. Dec 18, 2014. https://www.crowell.com/NewsEvents/AlertsNewsletters/all/New-York-Issues-Statewide-Fracking-Ban (accessed Oct 28, 2020).

33 Wu X, Braun D, Schwartz J, Kioumourtzoglou MA, Dominici F. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Sci Adv* 2020; **6:** eaba5692.

34 Callaway B, Sant'Anna PHC. Difference-in-differences with multiple time periods. *J Econom* 2021; **225:** 200–30.

35 Kropko J, Kubinec R. Interpretation and identification of within-unit and cross-sectional variation in panel data models. *PLoS One* 2020; **15:** e0231349.

36 Hill TD, Davis AP, Roos JM, French MT. Limitations of fixed-effects models for panel data. *Sociol Perspect* 2020; **63:** 357–69.

37 Dave D, Friedson AI, Hansen B, Sabia JJ. Association between statewide COVID-19 lottery announcements and vaccinations. *JAMA Health Forum* 2021; **2:** e213117.

38 Goodman-Bacon A. Difference-in-differences with variation in treatment timing. *J Econom* 2021; **225:** 254–77.

39 Sun L, Abraham S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *J Econom* 2021; **225:** 175–99.

40 Imai K, Kim IS. On the use of two-way fixed effects regression models for causal inference with panel data. *Polit Anal* 2021; **29:** 405–15.

41 Borusyak K, Jaravel X, Spiess J. Revisiting event study designs: robust and efficient estimation. *arXiv* 2021; published online Aug 27. https://doi.org/10.48550/arXiv.2108.12419 (preprint).

42 de Chaisemartin C, D'Haultfœuille X. Two-way fixed effects estimators with heterogeneous treatment effects. *Am Econ Rev* 2020; **110:** 2964–96.

43 Baker A, Larcker DF, Wang CCY. How much should we trust staggered difference-in-differences estimates? European Corporate Governance Institute—Finance Working Paper No. 736/2021. January, 2022. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3794018 (accessed Feb 12, 2023).

44 HEI Panel on the Health Effects of Traffic-Related Air Pollution. Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects. Boston, MA: Health Effects Institute; 2010. https://www.healtheffects.org/publication/traffic-related-air-pollution-critical-review-literature-emissions-exposure-and-health (accessed July 13, 2020).

45 Dominici F, Peng RD, Bell ML, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA* 2006; **295:** 1127–34.

46 Jo YS, Lim MN, Han YJ, Kim WJ. Epidemiological study of $PM_{2.5}$ and risk of COPD-related hospital visits in association with particle constituents in Chuncheon, Korea. *Int J Chron Obstruct Pulmon Dis* 2018; **13:** 299–307.

47 Cortez-Lugo M, Ramírez-Aguilar M, Pérez-Padilla R, Sansores-Martínez R, Ramírez-Venegas A, Barraza-Villarreal A. Effect of personal exposure to PM2.5 on respiratory health in a Mexican panel of patients with COPD. *Int J Environ Res Public Health* 2015; **12:** 10635–47.

48 Wei Y, Wang Y, Di Q, et al. Short term exposure to fine particulate matter and hospital admission risks and costs in the Medicare population: time stratified, case crossover study. *BMJ* 2019; **367:** l6258.

49 Chiu HF, Peng CY, Wu TN, Yang CY. Short-term effects of fine particulate air pollution on ischemic heart disease hospitalizations in Taipei: a case-crossover study. *Aerosol Air Qual Res* 2013; **13:** 1563–69.

50 Morris RD. Airborne particulates and hospital admissions for cardiovascular disease: a quantitative review of the evidence. *Environ Health Perspect* 2001; **109:** 495–500.

51 Leiva G MA, Santibañez DA, Ibarra E S, Matus C P, Seguel R. A five-year study of particulate matter (PM2.5) and cerebrovascular diseases. *Environ Pollut* 2013; **181:** 1–6.

52 Gutiérrez-Avila I, Rojas-Bracho L, Riojas-Rodríguez H, Kloog I, Just AC, Rothenberg SJ. Cardiovascular and cerebrovascular mortality associated with acute exposure to $PM_{2.5}$ in Mexico City. *Stroke* 2018; **49:** 1734–36.

53 Lin H, Guo Y, Di Q, et al. Ambient PM2.5 and stroke. *Stroke* 2017; **48:** 1191–97.

54 Department of Environmental Protection. Oil and gas reports. https://www.dep.pa.gov:443/DataandTools/Reports/Oil%20and%20Gas%20Reports/Pages/default.aspx (accessed July 22, 2019).

55 Adgate JL, Goldstein BD, McKenzie LM. Potential public health hazards, exposures and health effects from unconventional natural gas development. *Environ Sci Technol* 2014; **48:** 8307–20.

56 US Census Bureau. 2015 Cartographic Boundary File, 5-Digit ZIP Code Tabulation Area for United States, 1:500,000. https://catalog.data.gov/dataset/2015-cartographic-boundary-file-5-digit-zip-code-tabulation-area-for-united-states-1-500000 (accessed Feb 12, 2023).

57 National Center for Health Statistics. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). https://www.cdc.gov/nchs/icd/icd9cm.htm (accessed July 13, 2020).

58 Goeminne PC, Cox B, Finch S, et al. The impact of acute air pollution fluctuations on bronchiectasis pulmonary exacerbation: a case-crossover analysis. *Eur Respir J* 2018; **52:** 1702557.

59 Raji H, Riahi A, Borsi SH, et al. Acute effects of air pollution on hospital admissions for asthma, COPD, and bronchiectasis in Ahvaz, Iran. *Int J Chron Obstruct Pulmon Dis* 2020; **15:** 501–14.

60 Li R, Jiang N, Liu Q, et al. Impact of air pollutants on outpatient visits for acute respiratory outcomes. *Int J Environ Res Public Health* 2017; **14:** E47.

61 National Center for Health Statistics. International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). https://www.cdc.gov/nchs/icd/icd-10-cm.htm (accessed March 23, 2020).

62 Chernozhukov V, Kasahara H, Schrimpf P. The association of opening K-12 schools with the spread of COVID-19 in the United States: county-level panel data analysis. *Proc Natl Acad Sci USA* 2021; **118:** e2103420118.

63 O'Shaughnessy E, Barbose G, Wiser R, Forrester S, Darghouth N. The impact of policies and business models on income equity in rooftop solar adoption. *Nat Energy* 2021; **6:** 84–91.

64 Courtemanche CJ, Le AH, Yelowitz A, Zimmer R. School reopenings, mobility, and COVID-19 spread: evidence from Texas. National Bureau of Economic Research, 2021. https://doi.org/10.3386/w28753 (accessed Feb 12, 2023).

65 Alexander D, Karger E. Do stay-at-home orders cause people to stay at home? Effects of stay-at-home orders on consumer behavior. *Rev Econ Stat* 2021; published online Sept 22. https://doi.org/10.1162/rest_a_01108.

66 Ash E, Chen DL, Naidu S. Ideas have consequences: the impact of law and economics on American justice. National Bureau of Economic Research, 2022. https://doi.org/10.3386/w29788 (accessed Feb 12, 2023).

67 Ardiles LG, Tadano YS, Costa S, et al. Negative binomial regression model for analysis of the relationship between hospitalization and air pollution. *Atmos Pollut Res* 2018; **9:** 333–41.

68 Callaway B, Sant'Anna PHC. Pre-testing in a DiD Setup using the did package. https://bcallaway11.github.io/did/articles/pre-testing.html (accessed March 4, 2022).

69 New York State Department of Environmental Conservation. High-volume hydraulic fracturing in NYS. 2015. http://www.dec.ny.gov/energy/75370.html (accessed July 8, 2019).

70 Tabuchi H. Fracking firms fail, rewarding executives and raising climate fears. *The New York Times*. July 12, 2020. https://www.nytimes.com/2020/07/12/climate/oil-fracking-bankruptcy-methane-executive-pay.html (accessed July 15, 2020).

71 Hoek G, Krishnan RM, Beelen R, et al. Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environ Health* 2013; **12:** 43.

72 Bartik AW, Currie J, Greenstone M, Knittel CR. The local economic and welfare consequences of hydraulic fracturing. *Am Econ J Appl Econ* 2019; **11:** 105–55.

73 Brandt EJ, Myerson R, Perraillon MC, Polonsky TS. Hospital admissions for myocardial infarction and stroke before and after the trans-fatty acid restrictions in New York. *JAMA Cardiol* 2017; **2:** 627–34.