Received: 15 August 2016 | Revised: 11 December 2017 | Accepted: 28 January 2018

DOI: 10.1002/hec.3649

RESEARCH ARTICLE

WILEY Health Economics

The health implications of unconventional natural gas development in Pennsylvania

Lizhong Peng[1] | Chad Meyerhoefer[2] | Shin-Yi Chou[2]

[1]Department of Economics, University of West Georgia, Carrollton, GA, USA

[2]Department of Economics, Lehigh University and NBER, Bethlehem, PA, USA

Correspondence
Lizhong Peng, Department of Economics, University of West Georgia, 1601 Maple Street, Carrollton, GA 30118, USA.
Email: lpeng@westga.edu

Abstract

We investigate the health impacts of unconventional natural gas development of Marcellus shale in Pennsylvania between 2001 and 2013 by merging well permit data from the Pennsylvania Department of Environmental Protection with a database of all inpatient hospital admissions. After comparing changes in hospitalization rates over time for air pollution-sensitive diseases in counties with unconventional gas wells to changes in hospitalization rates in nonwell counties, we find a significant association between shale gas development and hospitalizations for pneumonia among the elderly, which is consistent with higher levels of air pollution resulting from unconventional natural gas development. We note that the lack of any detectable impact of shale gas development on younger populations may be due to unobserved factors contemporaneous with drilling, such as migration.

KEYWORDS

air pollution, fracking, pneumonia, shale gas development

1 | INTRODUCTION

Natural gas has become a key source of energy in the United States. Over the past decades, technological advancements in horizontal drilling and hydraulic fracturing (often referred to as "fracking") have made natural gas trapped beneath various shale formations more economically accessible. The contribution of shale gas to total U.S. natural gas production increased significantly from less than 2% in 2000 to over 20% in 2010; it is also projected that 46% of the natural gas supply will come from shale gas by 2035 (Energy Information Administration, 2013). With this rapid expansion of shale gas development, the potential health risks have drawn attention from the public and regulators at various levels.

Typically, the process to develop shale gas wells involves well pad preparation and construction, drilling and well construction, hydraulic fracturing, flaring of excess natural gas, and gas extraction and compression. Air pollution can occur during each stage of the process, whereas water contamination mostly occurs during wellbore drilling and hydraulic fracturing (see Appendix A.1 for a more detailed description of the stages of shale gas development). Numerous studies have documented that emissions of greenhouse gases (predominantly water vapor, carbon dioxide, methane, and ozone), volatile organic compounds (VOCs), other air pollutants, and hazardous chemicals increase as a result of unconventional natural gas development (Government Accountability Office (GAO), 2012; National Resources Defense Council, 2014). Based on data from a natural gas emissions inventory created by the Pennsylvania Department of Environmental Protection (PADEP) in 2011, levels of air pollutant

TABLE 1 Statewide air emissions from unconventional natural gas development in Pennsylvania, 2011–2013

Emissions (tons)	2011	2012	2013	Change 2011–2012 (%)	Change 2012–2013 (%)
Air pollutants					
CO	6,852	7,350	6,606	7.26	−10.12
NOx	16,542	16,361	17,659	−1.09	7.93
PM-10	577	600	670	4.05	11.67
PM-2.5	505	548	616	8.56	12.41
SOx	122	101	159	−17.32	57.43
VOC	2,820	4,024	4,790	42.69	19.04
Benzene	20	25	32.5	26.97	30.00
Ethylbenzene	5	6	11.5	11.25	91.67
Formaldehyde	251	374	404.4	48.93	8.13
n-Hexane	51	98	129.2	92.96	31.84
Toluene	34	33	43.5	−2.64	31.82
Xylenes	26	34	37.1	32.30	9.12
2,2,4-Trimethylpentane	4	19	18.9	439.11	−0.53
Greenhouse gases					
CO_2	N/A	4,291,316	4,908,106	N/A	14.37
Methane	N/A	123,884	107,945	N/A	−12.87
Nitrous oxide	N/A	209.3	77.8	N/A	−62.83
Cumulative number of unconventional wells since 2001	4,873	6,223	7,438	27.70	19.52

Note. These emission data are only for the drilling and production phases and do not include the emissions from the well-pad construction phase. Emissions for CO_2, methane, and nitrous oxide are not available for 2011. CO = carbon monoxide; NOx = nitrogen oxides; PM = particulate matters; VOC = volatile organic compound.

emissions (carbon monoxide, nitrogen oxides, $PM_{2.5}$, PM_{10}, etc.) attributable to unconventional natural gas drilling increased between 2011 and 2013 as the number of gas wells in the state rose by nearly 30% from 2011 to 2012 and 20% from 2012 to 2013 (see Table 1).[1]

Despite the fact that air pollution has clear adverse health effects, there is little scientific research on the impact of shale gas development on human health (see Appendix A.2 for a detailed discussion of the link between air pollution and human health). We know of only a few studies that investigate the direct impact of shale gas development on health. In particular, Hill (2012) finds a higher incidence of low birth weight among babies born to mothers living in the vicinity of shale gas wells in Pennsylvania. Casey et al. (2016) investigate the associations between unconventional natural gas development and birth outcomes using a linked dataset that contains information on both mothers and neonates. They find that exposure to unconventional natural gas development activities (measured by proximity to unconventional gas wells) is associated with increased risk of preterm birth. Jemielita et al. (2015) report that shale gas development between 2007 and 2011 in two counties (Bradford and Susquehanna) in Pennsylvania is significantly associated with increased hospitalizations for cardiology and neurology patients. They also find suggestive evidence of increased hospitalization rates for dermatology, oncology, and urology.

In this paper, we examine the impact of unconventional natural gas drilling in Pennsylvania on the treatment of several air-pollution-sensitive medical conditions at Pennsylvania hospitals. The state of Pennsylvania is rich in Marcellus shale reserves[2] and has witnessed a significant expansion of unconventional natural gas development in the past decade, making it a good location to study the effects of drilling. Our paper extends the work of Jemielita et al. (2015) in several important dimensions. First, we expand the scope of analysis from a few counties to the entire state of Pennsylvania in order to capture potential spill-over effects and increase the generalizability of our results. Second, the timeframe of our study is from 2001 to 2013, which encompasses the recent expansion in unconventional natural gas extraction. The longer time horizon allows us to better control for secular trends in hospitalization rates. The

[1]Pennsylvania Department of Environmental Protection (PADEP). Air Emissions Data from Natural Gas Operation. Accessed on June 24, 2017 at http://www.dep.pa.gov/business/air/baq/businesstopics/emission/pages/marcellus-inventory.aspx

[2]Marcellus shale (also known as the Marcellus formation) is a geological formation found in eastern North America, spanning six states in the northeastern United States. It is a unit of marine sedimentary rock that contains largely untapped natural gas reserves.

number of shale gas wells increased exponentially in 2012 and 2013, and incorporating these data allows us to uncover new health effects and assess the robustness of the results reported in Jemielita et al. (2015).[3] Third, we include a rich set of county characteristics in our fixed-effects regression models. This allows us to control for both time-varying observed and time-invariant unobserved confounders. Our identification strategy exploits changes in the timing of drilling in each county. We also use synthetic control methods to evaluate the robustness of our fixed-effects regression results.

We focus on the effect of Marcellus well development on the county-level hospitalization rates for acute myocardial infarction (AMI), chronic obstructive pulmonary disease (COPD), asthma, pneumonia, and upper respiratory infections (URI). We find that unconventional natural gas development was associated with a significant increase in the hospitalization rate for pneumonia among the elderly, which is consistent with higher levels of air pollution. However, results on other conditions are mixed and not robust to empirical methods and model specifications. Due to potential unobserved local demographic or economic changes contemporaneous with drilling, we caution that the magnitude of our estimates may not be as reliable as their directions.

2 | DATA

We obtained data on natural gas wells from the PADEP Oil and Gas Reports. The Spud Data Report contains information on the drilling commencement date (i.e., spud date), location, operator, and configuration of all conventional and unconventional natural gas wells drilled in Pennsylvania between 2001 and 2013. Unfortunately, these data do not include the well completion date, which indicates when the well is ready to produce natural gas.[4] We also obtained data on total annual gas production from the PADEP statewide well production database, which contains this information for all active natural gas wells. We linked spud dates to the gas production data using a unique well permit number, allowing us to determine the annual gas production for each active well after its spud date. Table 2 contains a list of all Pennsylvania counties that had unconventional natural gas wells drilled during the timeframe of this study. In particular, a county is classified as a "well county" if it has at least one unconventional wells drilled between 2001 and 2013, and as a "nonwell county" otherwise.

Our health outcome measures are derived from the Pennsylvania Health Care Cost Containment Council's (PHC4) compilation of all inpatient hospital admission records in the state during 2001–2013. We use International Classification of Diseases, Ninth Revision, Clinical Modification codes to identify the main diagnosis for each inpatient admission and then group related diagnoses into clinically meaningful categories using the Clinical Classification Software developed by the Agency for Healthcare Research and Quality (see Table A1 for how each condition is defined).[5] We focus our analysis on the following five health conditions that are sensitive to air pollution: AMI, COPD, asthma, pneumonia, and URI. Although many known or possible carcinogenic chemicals are used during the well development process (such as the benzene, toluene, ethylbenzene, and xylene [BTEX] compounds), we do not consider cancer in this analysis due to the short time frame of our data.[6] Persistent correlations between the above five conditions and air pollution have been established in the epidemiological literature (e.g., see Brunekreef and Holgate (2002); Dockery and Pope III (1994); Dominici, Peng, Bell, et al. (2006); Eder, Ege, and von Mutius (2006)); Garshick (2014); Lee, Kim, and Lee (2014); Neupane et al. (2010); Schwartz, Slater, Larson, Pierson, and Koenig (1993).

[3]Assessing the robustness of these estimates is particularly important given what appears to be an error in the dataset constructed by Jemielita et al. (2015). After aggregating zip code well counts to the county level from 2007 to 2011, the total number of unconventional wells drilled during this period in the Jemielita et al. (2015) dataset is 3,568 for Bradford County and 964 for Susquehanna County, respectively. However, according to the PADEP oil and gas reporting website (the source of raw well data for both Jemielita et al. (2015) and our study), there are records for only 945 unconventional wells drilled in Bradford county and 457 wells unconventional wells drilled in Susquehanna County between 2007 and 2011. The county-level well counts in our data are the same as those reported on the PADEP website, which offers interactive summary reports that contain the numbers of conventional and unconventional wells drilled in Pennsylvania counties. These reports can be accessed at: http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Oil_Gas/Wells_Drilled_By_County.

[4]For wells that begin producing immediately after drilling, the spud date and completion date are the same.

[5]The Clinical Classification Software can be accessed at http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.

[6]For example, a recent study measuring the impact of diesel exhaust exposure on the incidence of lung cancer among miners used a 15-year lag between the time of exposure and diagnosis of the illness (Attfield et al., 2012).

TABLE 2 Unconventional and conventional natural gas wells drilled in Pennsylvania between 2001 and 2013

Well counties ($N = 39$)				Nonwell counties ($N = 28$)	
County	**Drilling year**	**Unconventional wells**	**Conventional wells**	**County**	**Conventional wells**
Allegheny	2008	31	468	Adams	0
Armstrong	2006	178	2,929	Berks	0
Beaver	2009	35	4	Bucks	0
Bedford	2010	1	15	Carbon	0
Blair	2010	6	0	Chester	0
Bradford	2005	1,219	29	Cumberland	0
Butler	2006	268	119	Dauphin	0
Cambria	2009	7	180	Delaware	0
Cameron	2008	20	14	Erie	272
Centre	2007	63	372	Franklin	0
Clarion	2007	25	1,496	Fulton	0
Clearfield	2007	150	1,307	Juniata	0
Clinton	2008	100	102	Lancaster	0
Columbia	2010	3	0	Lebanon	0
Crawford	2012	3	1,414	Lehigh	0
Elk	2005	71	963	Mifflin	0
Fayette	2006	249	2,269	Monroe	0
Forest	2009	22	2,336	Montgomery	0
Greene	2006	631	934	Montour	0
Huntingdon	2010	1	4	Northampton	0
Indiana	2003	49	2,799	Northumberland	0
Jefferson	2008	40	1,861	Perry	0
Lackawanna	2009	2	0	Philadelphia	0
Lawrence	2011	30	30	Pike	0
Luzerne	2010	2	0	Schuylkill	0
Lycoming	2007	823	7	Snyder	0
Mckean	2006	73	5,934	Union	0
Mercer	2012	26	1,440	York	0
Potter	2007	70	336	N/A	N/A
Somerset	2004	26	19	N/A	N/A
Sullivian	2010	82	0	N/A	N/A
Susquehanna	2006	861	4	N/A	N/A
Tioga	2006	843	24	N/A	N/A
Venango	2011	6	1,157	N/A	N/A
Warren	2007	5	3,443	N/A	N/A
Washington	2002	974	661	N/A	N/A
Wayne	2008	5	4	N/A	N/A
Westmoreland	2003	258	2,447	N/A	N/A
Wyoming	2009	180	1	N/A	N/A
Total		7,438	35,122	N/A	272

Note. The numbers of conventional and unconventional wells represent the cumulative number of wells drilled for each type between the commencement date and 2013. Well counties are those that have at least one unconventional well during this period.

3 | EMPIRICAL METHODS

Our analysis sample consists of individuals aged 5 and above. We exclude young children aged 0–4 because about 78% of our hospital admission records in this age range are for newborns, and very young children are significantly less likely to be affected by air pollution than school-aged children due to the limited amount of time they spend outdoors. In addition to estimating models using the full sample, we stratify the sample into four age groups: 5–19, 20–44, 45–64, and above 65 because the pollution caused by shale gas development might have a more pronounced effect on one age group than others. For example, elderly individuals with a preexisting respiratory illness are more sensitive to air pollution and are more likely than younger individuals to contract pneumonia (Ryan et al., 2013). Because hospitalization rates for

AMI and COPD are either very low or zero among children ages 5–19, we only estimate models for asthma, pneumonia, and URI for this age group. In each year, we construct county-level measures of the five health conditions by first aggregating the individual-level PHC4 data into county-year cells, and then normalizing the total number of inpatient admissions for each condition by the population of that county in each age group in the same year,[7] so that the measures reflect hospitalizations per 1,000 people. We chose to conduct our analysis at the county level instead of zip code level as in Jemielita et al. (2015) mainly because annual population estimates and demographic and economic characteristics are only available at the county level from the Census Bureau. In particular, normalizing the county-level hospitalization counts and controlling for county-level demographic shifts, such as changes in the age distribution, in regression models should reduce the influence of population migration on our estimates.

3.1 | Difference-in-differences

The average hospitalization rates for AMI, COPD, asthma, pneumonia, and URI are higher in counties containing unconventional natural gas wells than those without wells. This is likely because poverty rates are 28% higher in well counties, on average, and the fact that these counties have much higher levels of extraction of other natural resources, such as coal (see Table 3). Furthermore, the hospitalization rates for certain conditions and age groups exhibit strong secular trends over time in both well and nonwell counties (see Figure A1). In order to address both of these issues, we estimate the impact of unconventional natural gas development on county-level hospitalization rates using a difference-in-differences strategy. Specifically, we exploit the plausibly exogenous variation in the timing of drilling in each county (see Table 2). We estimate the following "staggered treatment" specification:

$$\mathrm{y_{ct}} = \mathrm{X}_{\mathrm{ct}}'\beta + \sum_{\mathrm{r}=0}^{\mathrm{s}} \alpha_{\mathrm{r}} \mathrm{W}_{\mathrm{c,t-r}} + \sum_{\mathrm{r}=0}^{\mathrm{s}} \theta_{\mathrm{r}} \mathrm{L}_{\mathrm{c,t-r}} + \psi_{\mathrm{c}} + \zeta_{\mathrm{t}} + \mathrm{u_{ct}}, \tag{1}$$

where y_{ct} is the health outcome of interest in county c and year t, X_{ct} is a vector of county characteristics (more discussion below), W_{ct} is an indicator variable that equals to one if there are active unconventional wells in the county in year t,[8] L_{ct} is the log of natural gas output[9] from all active unconventional wells in the county in year (we add 1 to the level of output in counties without wells because output is equal to 0 for them), and is a random error term assumed to be uncorrelated with all of the regressors. The county fixed effects ψ_c and year fixed effects control for time-invariant unobserved county-level heterogeneity and overall common shocks to the outcomes, respectively. We also augment our baseline model with county-specific linear trends. In the baseline model, identification of the treatment effects comes from comparing changes in hospitalization rates over time in counties with unconventional wells relative to the change in hospitalization rates over time in counties without unconventional wells (with a similar number of counties in each group). However, identification in the more demanding models with county-specific linear trends comes from sharp deviations in the outcomes from linear trends in the treatment counties relative to the control counties. As a result, this specification is robust to preexisting differential trends in the outcomes across the treatment and control counties.

We use one lag for both the binary well indicator and log of output. The main reason for choosing this specification is that air pollution may have lagged effects on health, especially for some of the chronic conditions we examine. Even for acute conditions like asthma, it is possible that an individual must be exposed to certain air pollutants for an extended period of time before experiencing symptoms. We note that air pollution may come from gas well development activities (site construction, drilling, and initial hydraulic fracturing) as well as ongoing extraction activities (gas production, compression, and fuel transportation). The treatment dummies in (1) should capture the effects of well development activities, whereas the inclusion of the log of output in our model will capture the intensity of ongoing extraction activities. The parameters of interest, α and θ, measure the reduced-form effects of Marcellus shale gas development and production on the county-level hospitalization rate for each medical condition.

Our models also contain control variables for county-level demographic composition, economic conditions and activities, and measures of patient characteristics at each county's hospitals (see Table 3). Specifically, we include the county-level unemployment rate, poverty rate, the log of county population density, quartiles of county median household income, and the percentage of the county population in each 5-year age category, from 0 to 4 to 85 and above. These

[7]These data are obtained from the Census Bureau Population Estimates Program, which can be accessed at http://www.census.gov/popest/.

[8]Once an unconventional well is drilled, the indicator W_{ct} remains 1 for the rest of the sample period.

[9]Natural gas output is measured in thousands of cubic feet.

TABLE 3 Summary statistics, county-level, 2001–2013

	Full sample (Age 5 and above)		Age 5–19		Age 20–44		Age 45–64		Age 65 and above	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Patient characteristics										
Age	59.001	1.787	14.424	0.430	32.011	0.616	55.179	0.402	77.982	0.705
Female	0.576	0.019	0.570	0.046	0.701	0.033	0.495	0.024	0.562	0.026
White	0.916	0.097	0.847	0.142	0.881	0.121	0.909	0.102	0.947	0.079
Black	0.038	0.070	0.071	0.098	0.053	0.081	0.045	0.080	0.022	0.053
Asian	0.003	0.003	0.004	0.005	0.005	0.007	0.002	0.003	0.001	0.002
Other race	0.043	0.053	0.078	0.076	0.061	0.063	0.043	0.050	0.030	0.053
Hispanic	0.017	0.041	0.038	0.059	0.027	0.049	0.015	0.040	0.009	0.040
Private	0.324	0.049	0.589	0.088	0.545	0.079	0.574	0.082	0.060	0.028
Medicare	0.516	0.043	0.004	0.008	0.083	0.023	0.227	0.047	0.924	0.034
Medicaid	0.127	0.033	0.364	0.090	0.308	0.065	0.153	0.046	0.004	0.005
Government insurance	0.010	0.006	0.014	0.011	0.013	0.011	0.015	0.012	0.005	0.006
Self-pay	0.018	0.012	0.021	0.016	0.045	0.020	0.025	0.014	0.002	0.010
Other insurance	0.005	0.009	0.007	0.011	0.006	0.009	0.005	0.008	0.005	0.012
Admission: emergency	0.428	0.172	0.415	0.152	0.322	0.137	0.427	0.168	0.479	0.205
Admission: urgent	0.269	0.171	0.286	0.140	0.241	0.133	0.261	0.165	0.283	0.205
Admission: elective	0.299	0.062	0.293	0.111	0.431	0.098	0.308	0.058	0.235	0.058
Admission: other type	0.000	0.001	0.000	0.002	0.000	0.001	0.000	0.001	0.000	0.001
Charlson index	1.080	0.186	0.157	0.064	0.285	0.062	1.108	0.163	1.522	0.270
County-level hospitalization rate (cases per 1,000 people)										
Acute myocardial infarction (AMI)	3.307	0.995	0.002	0.011	0.447	0.243	3.577	1.090	12.080	3.720
Chronic obstructive pulmonary disease (COPD)	3.157	1.294	0.018	0.044	0.223	0.201	3.001	1.390	12.450	4.474
Asthma	1.144	0.617	0.719	0.749	0.780	0.449	1.296	0.791	2.075	1.226
Pneumonia	4.544	1.486	0.737	0.435	1.008	0.466	3.237	1.269	17.686	5.484
Upper respiratory infection (URI)	0.478	0.253	0.225	0.221	0.236	0.175	0.387	0.268	1.380	0.823

	Well counties ($N = 507$)		Nonwell counties ($N = 364$)	
County characteristics (across all age groups)	Mean	Standard deviation	Mean	Standard deviation
Poverty rate[a] (%)	13.351	2.719	10.401	3.759
Household median income[a] (thousand, in 2013 dollars)	44.515	4.452	56.191	11.573
Unemployment rate[a] (%)	6.832	1.775	6.133	1.936
Population density[a] (number of people per square mile)	4.520	1.019	5.744	1.209
Log of coal production (underground, short tons)[b]	3.521	6.107	1.506	3.189
Log of coal production (surface, short tons)[b]	7.508	6.061	0.906	3.233
Has an unconventional well in current year (0/1)[c]	0.491	0.500	N/A	N/A
Has an unconventional well in previous year (0/1)[c]	0.414	0.493	N/A	N/A
Log of unconventional output in current year[c] (million cubic feet)	5.128	6.976	N/A	N/A
Log of unconventional output in previous year[c] (million cubic feet)	4.069	6.409	N/A	N/A
Log of number of conventional wells[c]	2.318	2.222	0.103	0.522
Log of conventional output[c] (million cubic feet)	11.055	6.339	0.560	2.743

Note. Natural gas output is measured in thousand cubic feet (MCF); Household median income is unadjusted due to the methodological change of SAIPE after 2005. County-level incidences are calculated by first aggregating the individual-level PHC4 data to county-year cells for each age group, and then normalizing by the population in that age group. All patient characteristics are county-level means from the PHC4 data.

[a]U.S. Census Bureau, Small Area Income and Poverty Estimates (SAIPE). [b]U.S. Energy Information Administration. [c]Pennsylvania Department of Environmental Protection.

measures are derived from the Small Area Income and Poverty Estimates (SAIPE) and the Population Estimates Program data made available by the U.S. Census Bureau.[10,11]

Greater availability and lower prices of natural gas may reduce the demand for other fuels that generate pollution during their extraction and use (Environmental Protection Agency (EPA), 1999; National Research Council, 2010). As unconventional natural gas extraction in Pennsylvania has increased, the extraction of coal for electricity generation has decreased as has the extraction of oil and gas from conventional wells. In order to capture this substitution of fuels, we include in our models the log of annual county-level production of coal from surface and underground mining reported to the U.S. Energy Information Administration. We also include the annual number of new conventional wells and total output from all conventional wells (both on the log scale).[12]

Finally, we construct county-level measures of patient characteristics from the PHC4 inpatient admission records and include these in our models. These are the county-level proportion of female patients, the proportion of patients of different racial and ethnic categories (White, Black, Asian, Hispanic, and other races), the proportion of different types of admissions (elective, urgent, emergency, and other types), and the proportion of patients in mutually exclusive insurance categories (private insurance, Medicaid, Medicare, other government insurance, self-pay, and all other payers), and the county average Charlson index (Charlson, Pompei, Ales, & MacKenZie, 1987).[13] In Table 3, we report descriptive statistics for all variables used in this analysis by age group. As expected, the county-level hospitalization rates for pollution-sensitive conditions and the Charlson index increase with age.

For all of our regression models, we cluster the standard errors at county level to account for within-county correlations in the outcomes. However, one important consideration for this analysis is the multiple comparisons problem: It is possible that we find statistically significant estimates simply by chance because we are conducting a large number of hypothesis tests (23 in total). To address this issue, we calculate the family-wise error rate adjusted p values using the free step-down resampling method proposed by Westfall and Young (1993).[14] Compared with conventional methods such as the Bonferroni correction, one of the main features of this bootstrap-based approach is that it provides more statistical power by accounting for underlying correlations among the outcomes (Anderson, 2008; Kling et al., 2007). This is particularly attractive for our application because we expect strong intercorrelations among our outcomes (e.g., when we consider the same condition in different age groups). Some recent studies that employ the same method include Kling et al. (2007), Acemoglu and Finkelstein (2008), Gibson, McKenzie, and Stillman (2011), and Liebman and Luttmer (2015).

3.2 | Synthetic control method

One critical issue for our study is the timing of treatment. Even though we have information on the exact date when the first unconventional well was drilled in each treatment county, it is still unclear ex-ante how much exposure to unconventional wells constitutes a treatment sufficient to impact human health. In fact, the amount of pollution generated from initial well pad (site) preparation, well construction, and hydraulic fracturing may vary greatly across wells given the location and geological characteristics of the drilling site. In addition, there is a great deal of heterogeneity in terms of the trajectory of the number of unconventional wells among well counties. Importantly, the county-year panel structure of our data sample effectively limits the number of lags that can be included in these models.

In order to address these issues, we use the synthetic control method developed by Abadie, Diamond, and Hainmueller (2010) to complement the difference-in-differences models. One of the attractive features of the synthetic control method is that it allows us to examine the evolution of the impact of drilling for an extended period of time in a visually simple way. As in the difference-in-differences models, we code a county as treated during and after the year

[10]We use the annual quartiles of household income due the methodological changes in the SAIPE. For the series of SAIPE state and county estimates, there is a break between 2004 and 2005 due to the switch from the Current Population Survey Annual Social and Economic Supplement to the American Community Survey data in SAIPE modeling. For that reason, estimates for these particular years are not directly comparable.

[11]The Intercensal Population Estimates released by the U.S. Census Bureau provide estimates of county-level population by 5-year age groups. We include percentages of population in each age group for each county year in all of our models to account for the demographic changes in the counties.

[12]Because most oil and gas producing counties had conventional wells before the beginning of our sample period, we use the number of new conventional wells instead of a treatment dummy to better capture the change in the intensity of conventional well development.

[13]To avoid endogeneity, we only use contemporaneous secondary diagnoses when computing the Charlson index.

[14]We implemented this method using 10,000 bootstrap replications based on the modified algorithm in Kling, Liebman, and Katz (2007). See Section C of Online Appendix of Kling et al. (2007) for more details of the algorithm.

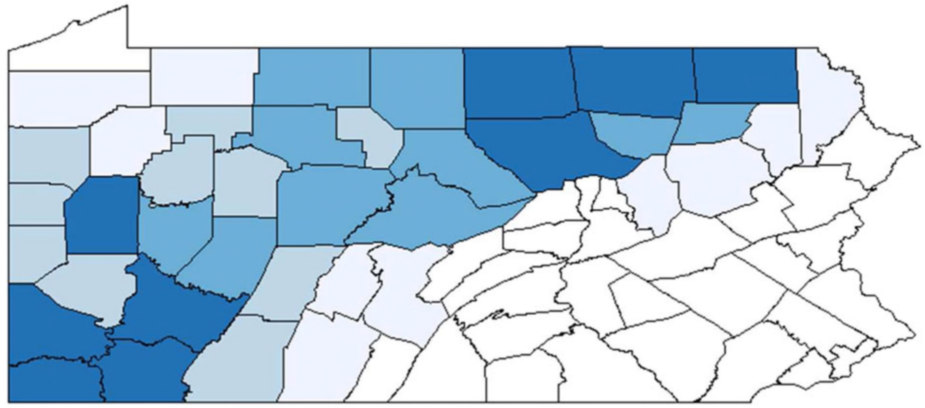


FIGURE 1 Pennsylvania counties with unconventional natural gas wells [Colour figure can be viewed at wileyonlinelibrary.com]

when the first unconventional well was drilled. We believe that this is the best approach given that activities surrounding the development of wells can generate significant pollution even before natural gas is extracted from the wells. We include in the control group the 28 counties that have no unconventional wells by the end of 2013. Operationally, for each well county, the synthetic control method selects a set of weights for the control counties so that the difference between the treatment and control counties in both the *outcome and observed characteristics* in the pretreatment period is minimized. The weighted average of the outcome for the control counties is then used as the counterfactual for the treated county in the post period. We are then able to calculate a treatment effect for each of the posttreatment years. We focus on the eight counties that are in the top quartile of the number of unconventional wells drilled by 2013 (Figure 1). We choose these counties because all of them had at least one unconventional well drilled before 2007, enabling us to look at a relatively long posttreatment period.[15] This is particularly appealing because unconventional natural gas development activities started to increase dramatically after 2008 and peaked around 2011 (see Figure 2 for the number of new unconventional wells drilled each year during our sample period). That is, we are able to see how the health effects (if there are any) vary during this period of rapid expansion of well development activities.

To streamline the presentation of the results, we use the extension proposed by Cavallo, Galiani, Noy, and Pantano (2013) to aggregate the treatment effects. In implementing the synthetic control method, we use the same control variables as in the difference-in-differences models (except for coal production and unconventional well counts and output) and the average outcome in the pretreatment period to construct the synthetic matches.[16] We use permutation tests (placebo tests) to conduct statistical inference (Abadie et al., 2010; Cavallo et al., 2013). Intuitively, the estimated effects are statistically significant when they are extreme relative to an empirical distribution of placebo effects (see Appendix A.4 for a detailed description of how we implement the synthetic control method and conduct statistical inference).

4 | RESULTS

Overall, we find consistent evidence that unconventional natural gas development is positively associated with higher rates of pneumonia among the elderly. The results for other conditions are somewhat mixed and sensitive to model specification. To simplify the presentation of our results, we discuss our findings on pneumonia in this section and relegate the results for AMI, COPD, asthma, and URI to the Appendix.

[15]These eight counties are Bradford, Butler, Fayette, Greene, Lycoming, Susquehanna, Tioga, and Westmoreland.The synthetic control method requires at least two pre-intervention years. Therefore, we did not include Washington County in this analysis even though it is in the top quartile because the first unconventional well was drilled in 2002 and our analysis sample starts in 2001.

[16]As can be seen in Table 3, control counties have minimal coal production and very few conventional wells compared with treatment counties. Therefore, we do not include these variables in the synthetic matching procedure to improve the match between the treatment counties and their synthetic counterparts.

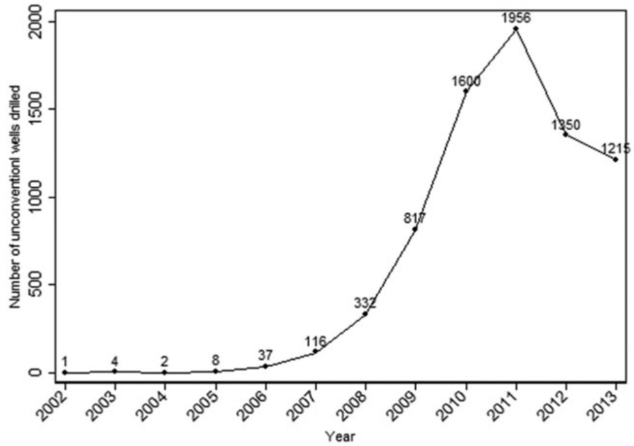


FIGURE 2 Number of new unconventional wells drilled each year between 2001 and 2013

4.1 | Difference-in-differences estimates

We report the difference-in-differences estimates for pneumonia in Table 4 (see Table A2 for the full set of estimates for other conditions). For each age group, the first column presents estimates from the baseline model with county and year fixed effects, and the second column presents estimates from the model that additionally contains county-specific linear time trends. In the full sample, the baseline model suggests that unconventional well development activities in the current and previous year are associated with an increase of 0.2 and 0.3 pneumonia admissions per 1,000 people, respectively (5.3% and 6.4% relative to the state mean). When we add county-specific linear trends to the model, the estimated effects remain positive but become smaller and lose statistical significance. This is somewhat expected as such trends may remove useful variation from identification and attenuate the estimates. In addition, we do not find any statistically significant impact of unconventional output on the pneumonia hospitalization rate in either specification.

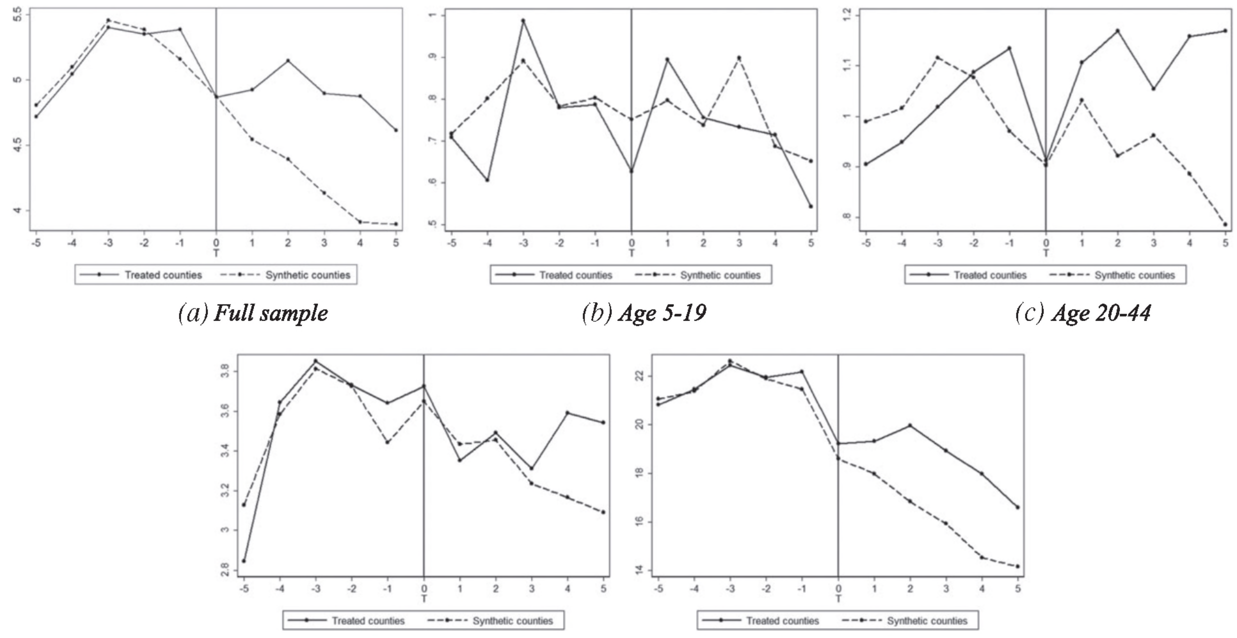
Columns 3 to 10 of Table 4 show the estimates by age group. In the baseline model, we find that unconventional well development in the previous year is associated with an increase of 1.5 admissions per 1,000 people, or 8.5% relative to the state mean, among the elderly. The effect remains statistically significant in the augmented model but decreases to about 1 admission per 1,000 people, or 5.7% relative to the state mean. However, we do not find the contemporaneous or lagged

TABLE 4 Impact of shale gas development on county-level hospitalization rate for pneumonia, 2001–2013

	Full sample (Age 5 and above)		Age 5–19		Age 20–44		Age 45–64		Age 65 and above	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Well current year	0.241*	0.149	0.062	0.098	0.066	0.026	0.141	0.019	0.777	0.506
	(0.131)	(0.133)	(0.073)	(0.080)	(0.056)	(0.061)	(0.152)	(0.151)	(0.479)	(0.570)
Well last year	0.294**	0.223	0.065	0.061	0.008	0.013	0.140	0.115	1.509***	0.995**
	(0.126)	(0.145)	(0.075)	(0.092)	(0.072)	(0.085)	(0.177)	(0.232)	(0.402)	(0.484)
Log output	0.005	0.002	−0.010	−0.007	0.001	0.000	−0.014	−0.020	0.000	0.012
	(0.018)	(0.017)	(0.008)	(0.008)	(0.007)	(0.008)	(0.017)	(0.023)	(0.067)	(0.065)
1st lag of log output	−0.000	−0.010	0.002	−0.003	0.009	0.005	0.007	−0.005	−0.037	−0.049
	(0.013)	(0.012)	(0.007)	(0.008)	(0.006)	(0.005)	(0.015)	(0.016)	(0.052)	(0.038)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County specific linear trends	N/A	Yes	N/A	Yes	N/A	Yes	N/A	Yes	N/A	Yes

Note. Standard errors are clustered at the county level. The total of number of observations is 804. All models include county and year fixed effects. Other control variables include average age, the share of different types of insurance (Medicare, Medicaid, private, self-pay, government, and other insurance), the share of female patients, the share of different race and ethnicity groups (White, Black, Asian, Hispanic, and other race), the share of different types of admission (emergency, urgent, elective, and other types), average Charlson index, county-level unemployment rate, poverty rate, annual quartiles of median household income, log of population density, log of annual coal production (both surface and underground), log of number of conventional wells, log of conventional output, and the entire county-level age distribution.

*$p < .1$. **$p < .05$. ***$p < .01$.



(a) Full sample (b) Age 5-19 (c) Age 20-44

(d) Age 45-64 (e) Age 65 and above

FIGURE 3 Evolution of the hospitalization rate for pneumonia in well counties in the top quartile of the well distribution and their synthetic matches

treatment dummy to be statistically significant in other age groups. Taken together, these results suggest that the adverse effects of unconventional well development on pneumonia are concentrated among the elderly.

Just as in the full sample, unconventional output does not seem to have any detectable effect on the hospitalization rate for pneumonia. This suggests that the health effects are mainly driven by unconventional well development activities instead of production-related activities. It is also worth noting that our measures of conventional drilling intensity (log of number of new conventional wells and log of conventional output) are statistically significant in only one of our 10 regressions on pneumonia.[17] This rules out the possibility that our estimated effects are due to changes in conventional well development and production activities.

To account for the multiple inference problem, we calculate the family-wise error rate adjusted p values using the method described above. As expected, the adjusted p values for some of our estimates become substantially larger. In particular, the contemporaneous and lagged treatment dummies in the full sample lose statistical significance after this adjustment (adjusted p values are 0.77 and 0.44, respectively). Among the elderly, the lagged treatment dummy remains statistically significant in the baseline model with an adjusted p value of 0.03, whereas the same effect drops below conventional levels of statistical significance in the augmented model (adjusted p value equals of 0.57).[18] Given that we are still able to detect an effect under this conservative approach, we conclude that it is unlikely that the relationship between drilling and pneumonia among the elderly that we find in multiple specifications is observed by chance.

4.2 | Synthetic control estimates

In Figure 3, we plot the evolution of the hospitalization rate for pneumonia averaged across the eight counties and their synthetic matches in the top quartile of the unconventional well distribution for the 5 years before and after the first unconventional well was drilled (see Figure A2 for other conditions). The full set of synthetic control estimates and root mean squared prediction error (RMSPE) in the pretreatment period are reported in Table A4. Due to the poor pretreatment fit for some of the outcomes (as indicated by the large RMSPE), we only conduct statistical inference for the cases where we have reasonably good pretreatment fit (RMSPE is less than 10% of the mean of the outcome). The poor fit of the synthetic matches for the rest of the cases also suggests the presence of unobserved factors that our models do not capture.

[17]The estimated coefficient on log of number of conventional wells is *negative* and statistically significant with an unadjusted p value of 0.07 in the augmented model for pneumonia among individuals aged 45–64.

[18]The full set of family-wise error rate adjusted p values is available from the authors upon request.

Panels A, D, and E of Figure 3 show that the synthetic matches provide a good approximation for the pretreatment trajectory of the pneumonia hospitalization rate in the full sample as well as for the two older age groups. The p values obtained from placebo tests indicate that the gap between treatment counties and their synthetic matches is not statistically significant among those aged 45–64 (see Panel D of Figure A3). On the other hand, we find that unconventional well development seems to have a long-lasting impact on the hospitalization rate for pneumonia among the elderly. As shown in Panel E of Figure A3, there is virtually no gap between the treatment counties and synthetic matches in the years before unconventional drilling occurred. However, the gap begins to widen after drilling started to intensify. Although not statistically significant, the treatment counties have 1.4 more admissions per 1,000 people 1 year after drilling began ($T = 1$) as compared with the synthetic matches. Interestingly, after the initial increase, the estimated effect tends to be constant and remain highly statistically significant; the average of second to fifth leads shows an increase of 2.7 admissions per 1,000 people in the treatment counties. We also find that the estimated effects exhibit the same pattern in the full sample (see Panel F of Figure A3). In general, these results lend support to our difference-in-differences estimates.

4.3 | Robustness and falsification tests

We conduct a series of additional checks in order to assess the robustness of our main results. Importantly, our estimates are reliable only if the model either captures or removes county-level factors that are correlated with both unconventional well development and hospitalization rates. We first subject the difference-in-differences models to several alternative specifications. If counties that are more likely to have higher growth in hospitalization rates are also more likely to have unconventional wells, then our estimates will be upwardly biased. In order to assess whether there exist such preexisting trends in the outcomes, we reestimate the baseline models with 2 leads of the treatment dummy (contemporaneous well development) and no output variables. We report the point estimates from this exercise in Table A3. Across the models for pneumonia, we do not find any statistically significant coefficient estimate on the lead treatment dummies. These results suggest that it is unlikely that our main finding on pneumonia is driven by preexisting differential trends within counties.

To assess whether our results are sensitive to the composition of the control group, we reestimate Equation 1 after excluding the five largest urban counties in Pennsylvania so that counties in the treatment and control groups are more comparable in terms of population density.[19] We find that our estimates are largely unchanged using this set of control counties. In another set of models, we test whether our results are robust to alternative treatment definitions. Specifically, we reestimate Equation 1 with treatment defined as (a) having at least five unconventional wells drilled to date or (b) having at least 100 unconventional wells drilled to date. Results from these models (not shown) indicate that the estimated effects on pneumonia become statistically insignificant under these alternative treatment definitions.[20] One likely explanation for this result is that the health effects of unconventional natural gas drilling are immediate (occur within one or 2 years after drilling begins) and do not increase proportionally with the number of wells, which is supported by the synthetic control estimates for pneumonia (Panels A and E of Figure 3). As a result, the difference-in-differences models with treatment delayed a few years may mask the true effects of drilling because the increase in hospitalization rates following the initial drilling years are now included in the "pretreatment" period under these alternative treatment definitions.[21] We provide a more detailed discussion of these results in Section 5.

Finally, we conduct a Monte Carlo simulation designed to detect spurious correlations between the well development indicators and outcome variables. We use the period of 2000–2005 and randomly assign placebo treatments to counties that contained wells in the post-2005 period by drawing the year of drilling in the prewell period from a uniform distribution. We then estimate Equation 1 for each condition with current and lagged placebo treatment indicators but without the output variables (see Appendix A.3 for a detailed description). In general, the probability of a Type I error (incorrectly rejecting the null hypothesis of no effect) is very close to the levels of significance we report for pneumonia (see Table A5). Specifically, we find that the coefficient estimate on the first lag of treatment dummy is significant at 1%

[19] These five counties are Philadelphia county (City of Philadelphia), Allegheny county (City of Pittsburgh), Lehigh county (City of Allentown), Erie county (City of Erie), and Berks county (City of Reading).

[20] These models include a treatment dummy and its first lag but do not include unconventional output. Results are available from the authors upon request.

[21] When 5 or 100 wells are used to define treatment, the well development indicator W_{ct} is set equal to 1 a few years later as compared with the original definition where W_{ct} is set to 1 once a well is drilled. If the health effects of drilling are immediate and relatively constant over time, then the difference-in-differences specification would fail to identify any significant effects under this alternative treatment definition because the years during which 1–5 or 1–100 wells were drilled would be included in the preperiod.

TABLE 5 Associations between county-level air emissions and unconventional natural gas development, 2011–2013

	(1) CO	(2) NOx	(3) PM-10	(4) PM-2.5	(5) SOx	(6) VOC	(7) Benzene
Log of number of new unconventional wells	0.417*** (0.128)	0.624*** (0.141)	0.351*** (0.067)	0.325*** (0.061)	0.159** (0.068)	0.314** (0.140)	0.021 (0.022)
Log of unconventional output	−0.059* (0.032)	−0.030 (0.037)	−0.012 (0.015)	−0.018 (0.013)	0.012 (0.013)	−0.002 (0.039)	−0.009* (0.004)
Log of number of conventional wells	−0.196 (0.139)	−0.201 (0.150)	−0.107 (0.086)	−0.128 (0.081)	−0.091 (0.110)	−0.022 (0.165)	−0.012 (0.032)
Log of conventional output	−0.092** (0.039)	−0.084* (0.045)	−0.019 (0.024)	−0.030 (0.019)	−0.165*** (0.051)	−0.142*** (0.041)	0.011 (0.010)

	(8) Ethyl Benzene	(9) Formaldehyde	(10) n-Hexane	(11) Toluene	(12) Xylenes	(13) 2,2,4-Trimethylpentane
Log of number of new unconventional wells	−0.005 (0.025)	0.032 (0.078)	0.007 (0.049)	0.021 (0.033)	0.056 (0.038)	−0.021 (0.035)
Log of unconventional output	−0.008 (0.006)	−0.008 (0.016)	−0.015 (0.009)	−0.006 (0.008)	−0.004 (0.008)	−0.015** (0.007)
Log of number of conventional wells	0.001 (0.039)	0.050 (0.099)	−0.054 (0.059)	−0.004 (0.053)	−0.028 (0.059)	0.043 (0.079)
Log of conventional output	0.006 (0.024)	0.025 (0.029)	−0.105*** (0.020)	0.034* (0.017)	−0.099** (0.040)	0.010 (0.041)

Note. Standard errors are clustered at the county level. The total number of observations is 105. 35 well counties are included in the regressions (Lackawanna and Luzerne do not have data for all three years, and Bedford and Wayne counties are not included in these data). Dependent variables include the log of total county-level emissions for the air pollutants. All models control for county and year fixed effects, the poverty rate, unemployment rate, population density, household median income, and coal production.

*$p < .1$. **$p < .05$. ***$p < .01$.

level for pneumonia among the elderly. And the simulations indicate that among 10,000 replications of a placebo treatment there are 125 times when the null hypothesis is falsely rejected, implying the probability of Type I error is 1.25%. In other words, the likelihood of a false positive in our models is nearly the same as a model that conforms perfectly to the assumptions of linear regression.

5 | DISCUSSION AND CONCLUSIONS

We examine the impact of unconventional natural gas development, which is known to cause air pollution, on human health. Our results show that horizontal drilling and hydraulic fracturing into Marcellus shale in the state of Pennsylvania over the past decade is associated with significant increases in hospitalization rates for pneumonia (among individuals aged 65 and above). Although we also find associations between natural gas development and extraction and AMI, COPD, asthma, and URI, these effects are sensitive to the empirical method as well as the functional specification of the models. Notably, we do not find any impact of gas well development on asthma, pneumonia, or URI among children aged 5–19. Because children spend more time outdoors, breath more rapidly than adults, and breath through their mouths rather than filtering air through their nose, their exposure to air pollution is typically assumed to be higher than adults (California Office of Environmental Health Hazard Assessment and the American Lung Association, 2003). It is possible that the impact of air pollution from well development has a longer term impact on children through the development of respiratory and other illnesses that we are unable to detect during the limited timeframe of our analysis. In contrast, the effects we find among adults may reflect the acute aggravation of preexisting conditions.

Another noteworthy finding is that most of the adverse effects we identify are likely due to well development and not natural gas extraction and compression, given that we do not find any unconventional output variables to be statistically significant in our models on pneumonia. Although we find positive and statistically significant associations between lagged gas output and AMI hospitalizations among middle-aged and elderly individuals (45-to 64-year olds and 65 and above), these output effects are not precisely estimated when county-specific linear trends are added to the difference-in-difference models (see Table A2). A significant source of air pollution during well development is the diesel engines that power heavy equipment used to build roads, clear well sites, construct wells, drill, and inject fracking fluid into the wells. Horizontal drilling followed by hydraulic fracturing is more energy-intensive than traditional vertical drilling, and the diesel engines used to pump fracking fluid commonly exceed 2,000 bhp (Treida, 2010).

Because our effects are predicated on increased exposure to air pollutants, it would be attractive to show a link between elevated levels of ambient air pollutant concentration and unconventional well development activities. Unfortunately, due to the poor overlap between air quality monitoring sites and well locations, we are not able to establish this relationship directly. That said, it is possible to use the PADEP county-level air emissions inventory (based on which we constructed Table 1) to examine if counties with more intensive well development activities generate more emissions. Because these data are only available starting in 2011, we compile a 3-year (2011–2013) county-level panel of air emissions and matched it to our well dataset. We then estimate a set of county fixed effects models that estimate the relationship between emissions and drilling intensity.[22] The key independent variables are the number of new unconventional wells each year and output from unconventional wells (both on the log scale). We also include the log of conventional well counts and output. The dependent variables are the log of annual emissions for a select group of air pollutants. We report the estimates from these regressions in Table 5. In general, we find strong and positive associations between the number of unconventional wells and county-level emissions of carbon monoxide, nitrogen oxides, PM_{10}, $PM_{2.5}$, sulfur oxides, and VOC. Despite these findings, it is important to point out that we are not able to assess whether the increases in air pollution emissions are large enough to affect human health due to a lack of air quality concentration measures.

Our analysis does have some limitations that should be kept in mind when interpreting the results. The large-scale development of Marcellus shale has inevitably caused migration into affected communities by gas workers and other individuals seeking jobs created by greater economic activity associated with growth in the gas industry. Likewise, shale gas development has resulted in out-migration by individuals that have sold their land to gas companies or have been displaced by the rising cost of housing in well counties. As shown in Panel A of Figure A4, the proportion of working age adults (aged 20–64) increased more quickly between 2008 and 2011 in well counties as compared with nonwell counties. This is consistent with the timing of increased drilling activities in the state. Although we have included

[22] In these models, we control for county and year fixed effects, the poverty rate, unemployment rate, population density, household median income, and coal production, all at the county-level.

variables in our empirical models that capture changes in the county-level age distribution over time, these variables do not capture the changes in underlying population health. Therefore, if the net impact of migration was to increase (decrease) the number of individuals with pollution-sensitive diseases in well counties, or decrease (increase) the number of individuals with these conditions in nonwell counties, our estimates will be upwardly (downwardly) biased. This raises the possibility that the lack of effects among younger individuals may be in part due to unobserved changes in county demographic characteristics. That said, our estimates on pneumonia among the elderly should be less affected by migration because there is no noticeable difference in the trend of proportion of individuals aged 65 and above between well and nonwell counties during the same time period (Panel B of Figure A4).

We find it interesting that the effects on pneumonia among the elderly are stable for the second to fourth leads even though drilling intensity increased substantially 3 or 4 years after drilling began (Panel E of Figure 3). One possible explanation for this is the diminishing amount of pollution caused by new wells. In fact, the horizontal drilling technique allows natural gas companies to drill multiple wells on a single well pad ("multiple-well pads"). Therefore, the pollution generated by new wells does not necessarily increase proportionally and the health effects may eventually level off, particularly if a large share of the pollution generated by wells occurs during the construction of the well pad. Another possibility is that higher concentrations of wells do pose greater public health risks, but there are unobserved county-level changes that occurred *contemporaneously* with drilling (which may or may not be related to unconventional natural gas development itself) that are not captured by our models. If these unobserved changes lowered the hospitalization rate, we could observe a relatively constant effect of drilling even though the real effects escalated over time. Unfortunately, we are not able to provide any further evidence to differentiate these explanations. As a result, we caution that our models may under-estimate the impact of natural gas development on pneumonia.

Despite these limitations, our study helps establish a consistent link between unconventional natural gas extraction and higher rates of disease. Our results have important implications for public policy because they provide evidence of an adverse impact of shale gas development on health, which is currently of concern to policy makers. For example, in April 2012, the U.S. EPA, Department of Energy, and Department of Interior agreed to collaborate on research in order to improve the "scientific understanding of hydraulic fracturing". In June 2015, the EPA released the results from a national study that investigates the potential impact of hydraulic fracturing on drinking water resources. They did not find evidence that hydraulic fracturing has led to widespread, systemic impacts on drinking water resources in the United States, despite some isolated cases of contaminated drinking water wells (EPA, 2015).

In 2010 and 2011, the Pennsylvania Department of Environmental Protection conducted three short-term studies to determine whether shale gas development affects air quality in the southwestern, northeastern, and northcentral regions of the state. In all three studies, natural gas constituents and associated compounds were detected in the air near Marcellus shale drilling operations, but the Department of Environmental Protection concluded that none of the compounds reached a level of concentration that could cause air-related health issues (Pennsylvania Department of Environmental Protection (PA DEP), 2010, 2011a, 2011b). However, a more recent study conducted by the Southwest Pennsylvania Environmental Health Project, a nonprofit environmental health organization, found short-term high values of particulate matter in the air and concluded that current methods of collecting and analyzing air pollutants emission data are not sufficiently accurate to evaluate the health risks of unconventional natural gas development (Brown, Weinberger, Lewis, & Bonaparte, 2014).

In the absence of strong scientific evidence on the relationship between shale gas development and health, states in the mid-Atlantic region have demonstrated conflicting regulatory objectives. In February 2012, the Pennsylvania General Assembly passed Act 13, which was a major overhaul of the state's oil and gas law. According to the new law, municipal governments are not allowed to impose stricter regulations on drilling activities than other industries and must allow oil and gas operations in "all zoning districts" (The General Assembly of Pennsylvania, 2011). This portion of the law resulted in disputes between local communities and state government and was subsequently ruled unconstitutional by the Pennsylvania Supreme Court (Cusick, 2013). In contrast, after conducting a 7-year study, the state of New York officially banned natural gas extraction activities that involve high-volume hydraulic fracturing on June 29, 2015 (New York Department of Environmental Conservation, 2015).

We seek to inform regulatory policies such as these by providing evidence on the link between Marcellus shale gas development and health. Because we find that unconventional natural gas well development has a stronger link to poor health than postdevelopment gas production, there is more limited justification for natural gas extraction taxes based on pollution-related externalities than per-well fees. The latter are similar, for example, to the annual "impact fees" levied on every Marcellus well drilled by operators under Pennsylvania's Act 13 of 2012. However, our results demonstrate a clear need for additional studies to confirm the precise causal pathways between unconventional gas well development, elevated levels of air pollution, and adverse health effects among different age groups.

DISCLAIMER

Pennsylvania inpatient data are from the Pennsylvania Health Care Cost Containment Council (PHC4). PHC4 is an independent state agency responsible for addressing the problem of escalating health costs, ensuring the quality of health care, and increasing access to health care for all citizens regardless of ability to pay. PHC4 has provided data to this entity in an effort to further PHC4's mission of educating the public and containing health-care costs in Pennsylvania. PHC4, its agents, and staff have made no representation, guarantee, or warranty and express or implied that the data —financial, patient, payer, and physician specific information—provided to this entity, are error-free or that the use of the data will avoid differences of opinion or interpretation. This analysis was not prepared by PHC4. PHC4, its agents, and staff bear no responsibility or liability for the results of the analysis. The authors have no conflict of interest.

ORCID

Lizhong Peng http://orcid.org/0000-0002-2680-4811

REFERENCES

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505.

Acemoglu, D., & Finkelstein, A. (2008). Input and technology choices in regulated industries: Evidence from the health care sector. *Journal of Political Economy*, *116*(5), 837–880.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495.

Attfield, M. D., Schleiff, P. L., Lubin, J. H., Blair, A., Stewart, P. A., Vermeulen, R., … Silverman, D. T. (2012 Jun 6). The diesel exhaust in miners study: A cohort mortality study with emphasis on lung cancer. *Journal of the National Cancer Institute*, *104*(11), 869–883.

Brown, D., Weinberger, B., Lewis, C., & Bonaparte, H. (2014). Understanding exposure from natural gas drilling puts current air standards to the test. *Reviews on Environmental Health*, *29*, 277–292.

Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The Lancet*, *360*(9341), 1233–1242.

California Office of Environmental Health Hazard Assessment and the American Lung Association. (2003). Air pollution and children's health. Accessed on March 19, 2015 at http://oehha.ca.gov/public_info/facts/airkids.html

Casey, J. A., Savitz, D. A., Rasmussen, S. G., Ogburn, E. L., Pollak, J., Mercer, D. G., & Schwartz, B. S. (2016). Unconventional natural gas development and birth outcomes in Pennsylvania, USA. *Epidemiology (Cambridge, Mass.)*, *27*(2), 163.

Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, *95*(5), 1549–1561.

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenZie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, *40*(5), 373–383.

Colborn, T., Kwiatkowski, C., Schultz, K., & Bachran, M. (2011). Natural gas operations from a public health perspective. *Human and Ecological Risk Assessment*, *17*(5), 1039–1056.

Committee on Energy and Commerce, U.S. House of Representatives. (2011). Chemicals Used in Hydraulic Fracturing. Accessed on May 5, 2014 at http://democrats.energycommerce.house.gov/sites/default/files/documents/Hydraulic-Fracturing-Chemicals-2011-4-18.pdf

Cusick M. (2013). Pennsylvania Supreme Court strikes down controversial portions of Act 13. Stateimpact Pennsylvania. Accessed on August 6, 2014 at http://stateimpact.npr.org/pennsylvania/2013/12/19/state-supreme-court-strikes-down-act-13-local-zoning-restrictions/

Department of Energy (2009). Modern shale gas development in the United States: A primer. Accessed on July 6, 2014 at http://www.netl.doe.gov/technologies/oil-gas/publications/EPreports/Shale_Gas_Primer_2009.pdf

Dockery, D. W., & Pope, C. A. III (1994). Acute respiratory effects of particulate air pollution. *Annual Review of Public Health*, *15*, 107–132.

Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L., & Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association*, *295*(10), 1127–1134.

Eder, W., Ege, M. J., & von Mutius, E. (2006). The asthma epidemic. *New England Journal of Medicine*, *355*(21), 2226–2235.

Energy Information Administration. (2013). Annual energy outlook 2013 with projections to 2040. Accessed on July 6, 2014 at http://www.eia.gov/forecasts/aeo/pdf/0383(2013).pdf

Environmental Protection Agency. (1998). NO_X: How nitrogen oxides affect the way we live and breathe. Accessed on May 5, 2014 at http://www.nchh.org/Portals/0/Contents/EPA%20Nitrogen_Oxides.pdf

Environmental Protection Agency. (1999). The benefits and costs of the clean air Act 1990–2010. Accessed on March 10, 2015 at http://www.epa.gov/oar/sect812/1990-2010/chap1130.pdf

Environmental Protection Agency. (2003). Particle pollution and your health. Accessed on May 5, 2014 at http://www.epa.gov/airnow/particle/pm-color.pdf

Environmental Protection Agency. (2009). Ozone and your health. Accessed on May 5, 2014 at http://www.epa.gov/airnow/ozone-c.pdf

Environmental Protection Agency. (2012). Well completions and workovers with hydraulic fracturing. Accessed on March 10, 2015 at http://www.epa.gov/climatechange/Downloads/ghgemissions/2012Workshop/EPA-Well_Completions_and_Workovers_with_Hydraulic_Fracturing.pdf

Environmental Protection Agency. (2015). Assessment of the potential impacts of hydraulic fracturing for oil and gas on drinking water resources: Executive summary. Accessed on March 28 at https://www.epa.gov/sites/production/files/2015-06/documents/hf_es_erd_jun2015.pdf

Garshick, E. (2014 Sep). Effects of short- and long-term exposures to ambient air pollution on COPD. *The European Respiratory Journal*, *44*(3), 558–561.

Gibson, J., McKenzie, D., & Stillman, S. (2011). The impacts of international migration on remaining household members: omnibus results from a migration lottery program. *Review of Economics and Statistics*, *93*(4), 1297–1318.

Government Accountability Office. (2012). Oil and gas: Information on shale resources, development, and environmental and public health risks. Accessed on July 6, 2014 at http://www.gao.gov/products/GAO-12-732

Hill, E. L. (2012). Shale gas development and infant health: Evidence from Pennsylvania. Cornell Dyson School working paper.

Jemielita, T., Gerton, G. L., Neidell, M., Chillrud, S., Yan, B., Stute, M., … Panettieri, R. A. Jr. (2015). Unconventional gas and oil drilling is associated with increased hospital utilization rates. *PLoS One*, *10*(7), e0131093.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.

Ko, F. W. S., Tam, W., Wong, T. W., Lai, C. K. W., Wong, G. W. K., Leung, T., … Hui, D. S. (2007). Effects of air pollution on asthma hospitalization rates in different age groups in Hong Kong. *Clinical & Experimental Allergy*, *37*, 1312–1319.

Lee, B. J., Kim, B., & Lee, K. (2014). Air pollution exposure and cardiovascular disease. *Toxicology Research*, *30*(2), 71–75.

Liebman, J. B., & Luttmer, E. F. (2015). Would people behave differently if they better understood social security? Evidence from a field experiment. *American Economic Journal: Economic Policy*, *7*(1), 275–299.

National Research Council. (2010). Hidden cost of energy: Unpriced consequences of energy production and use. Accessed on March 10, 2015 at http://www.nap.edu/catalog/12794/hidden-costs-of-energy-unpriced-consequences-of-energy-production-and

National Resources Defense Council. (2014). Fracking fumes: Air pollution from hydraulic fracturing threatens public health and communities. Accessed on March 10, 2015 at http://www.nrdc.org/health/files/fracking-air-pollution-IB.pdf

Neupane, B., Jerrett, M., Burnett, R. T., Marrie, T., Arain, A., & Loeb, M. (2010). Long-term exposure to ambient air pollution and risk of hospitalization with community-acquired pneumonia in older adults. *American Journal of Respiratory and Critical Care Medicine*, *181*, 47–53.

New York Department of Environmental Conservation. (2015). Final supplemental generic environmental impact statement on the oil, gas and solution mining regulatory program.

Pennsylvania Department of Environmental Protection. (2010). Southwestern Pennsylvania Marcellus shale short-term ambient air sampling report. Accessed on June 6, 2014 at http://www.dep.state.pa.us/dep/deputate/airwaste/aq/aqm/docs/Marcellus_SW_11-01-10.pdf

Pennsylvania Department of Environmental Protection. (2011a). Northcentral Pennsylvania Marcellus shale short-term ambient air sampling report. Accessed on June 6, 2014 at http://www.dep.state.pa.us/dep/deputate/airwaste/aq/aqm/docs/Marcellus_NC_05-06-11.pdf

Pennsylvania Department of Environmental Protection (2011b). Northeastern Pennsylvania Marcellus shale short-term ambient air sampling report. Accessed on June 6, 2014 at http://www.dep.state.pa.us/dep/deputate/airwaste/aq/aqm/docs/Marcellus_NE_01-12-11.pdf

Ryan, M., Suaya, J. A., Chapman, J. D., Stason, W. B., Shepard, D. S., & Thomas, C. P. (2013). Incidence and cost of pneumonia in older adults with COPD in the United States. *PLoS One*, *8*, e75887.

Schwartz, J., Slater, D., Larson, T. V., Pierson, W. E., & Koenig, J. Q. (1993). Particulate air pollution and hospital emergency room visits for asthma in Seattle. *American Review of Respiratory Disease*, *147*(4), 826–831.

The General Assembly of Pennsylvania. (2011). House bill No.1950, session of. Accessed on August 6, 2014 at http://www.legis.state.pa.us/WU01/LI/LI/US/HTM/2012/0/0013..HTM

Treida, M. (2010). The evolution of hydraulic fracturing and its effect on frac pump technology. Pumps & Systems. Accessed on March 19, 2015 at http://www.pumpsandsystems.com/topics/upstream-pumping/evolution-hydraulic-fracturing-and-its-effect-frac-pump-technology?page=2

Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment* (Vol. 279). New York, NY: John Wiley & Sons.

Zanobetti, A., & Schwartz, J. (2006). Air pollution and emergency admissions in Boston, MA. *Journal of Epidemiology & Community Health*, *60*, 890–895.

How to cite this article: Peng L, Meyerhoefer C, Chou S-Y. The health implications of unconventional natural gas development in Pennsylvania. *Health Economics*. 2018;27:956–983. https://doi.org/10.1002/hec.3649

APPENDIX A

A.1 | Shale gas development process

Unlike conventional natural gas development, extracting natural gas from unconventional formations relies heavily on horizontal drilling and hydraulic fracturing. Typically, operators first construct a well pad at the location suitable for drilling, build infrastructure, and transport equipment to the drilling site. In the next stage, a hole (wellbore) is drilled into the earth through a combination of vertical and horizontal drilling. Casing[23] and cement are inserted into the wellbore in order to isolate it from the surrounding formation. Finally, hydraulic fracturing is used to stimulate the shale formation. This involves the injection of highly pressurized fracturing fluid through the holes created by a perforating tool inserted in the casing and cement. As fracturing fluid is forced into the surrounding formation, fractures or cracks are created or expanded in the target formation. The underlying gas is then released and collected (GAO, 2012). It is worth noting that throughout the production period it may be necessary to restimulate the wells (also known as refracturing or well workovers) by repeating the hydraulic fracturing process, the frequency of which depends on the characteristics of geologic formation and production phase of a particular well (Department of Energy, 2009). When estimating annual greenhouse gas emission from natural gas production, the Environmental Protection Agency uses the assumption that 10% of unconventional wells need restimulation every year (EPA, 2012).

A.2 | Potential public health risks

Shale gas development and production may pose a threat to public health through air pollution (National Resources Defense Council, 2014). First, the construction of infrastructure at the drilling site requires massive transportation of water, sand, chemicals, and heavy machinery. Air pollutants such as NO_X and particulate matters (PM) contained in the engine exhaust brought about by increased traffic are released into the atmosphere. In addition, the development and production process requires a substantial amount of power, which is often supplied by diesel engines. The burning of diesel fuel also generates exhaust. Second, for operational reasons, flaring (burning) or venting (direct release into the atmosphere) of natural gas during the development and production process is sometimes necessary, which leads to emissions of carbon dioxide and the release of methane and VOCs. Third, evaporation of fracturing fluid and produced water may also emit hazardous chemicals into the atmosphere. Some of the air pollutants and chemicals from the drilling and gas production activities may be harmful to human health and even carcinogenic. NO_X can form small particles through reactions with ammonia, moisture, and other compounds. These particles penetrate deeply into the sensitive part of lungs and cause or worsen respiratory diseases (EPA, 1998). In addition, when reacting with VOCs in the presence of heat and sunlight, NO_X can form ground-level Ozone (smog), which irritates the respiratory system, reduces lung function, aggravates chronic conditions such as asthma and chronic bronchitis, and potentially results in permanent lung damage (EPA, 2009).

PM is also harmful. Short-term exposure to fine particles can cause asthma attacks and acute bronchitis and increases the risk of heart attacks and arrhythmias among people with heart disease (EPA, 2003). There are a multitude of studies that attempt to uncover the link between air pollution and adverse health outcomes. In general, researchers have found consistent evidence that air pollution is associated with respiratory problems. For example, Ko et al. (2007) find that levels of major air pollutants (NO_2, O_3, PM_{10}, and $PM_{2.5}$) in Hong Kong were associated with increased hospital admissions, with O_3 being the most important contributor. Likewise, Zanobetti and Schwartz (2006) find that air pollution in the greater Boston area was associated with a higher risk of hospitalization for pneumonia among individuals aged 65 and older.

Colborn, Kwiatkowski, Schultz, and Bachran (2011) compile a list of 632 chemicals used during the fracturing and drilling stages of natural gas development and report that many of them could have a negative impact on human health. In particular, more than 75% of the chemicals could affect the respiratory system; about half could affect the immune and cardiovascular systems; and 25% could cause cancer. A similar analysis was conducted in a congressional report by the Committee on Energy and Commerce of the U.S. House of Representatives. The report reviews the type and volume of hydraulic fracturing products used by 14 leading oil and gas companies between 2005 and 2009 and finds that the most widely used chemical during that period was methanol, which is a hazardous air pollutant, and that more than 650 hydraulic fracturing products contain 29 chemicals that are known or possible human carcinogens (Committee on Energy and Commerce, U.S. House of Representatives, 2011). These chemicals are either regulated under the Safe

[23]Casing is a metal pipe inserted inside the wellbore. It prevents fluids outside the formation from entering the well. It also protects fragile sections of the wellbore from drilling mud inside the well (GAO, 2012).

Drinking Water Act for their risks to human health and/or listed as hazardous air pollutants under the Clean Air Act. For instance, the BTEX compounds were found in many of the hydraulic products. Each BTEX compound is a regulated contaminant under the Safe Drinking Water Act and a hazardous air pollutant under the Clean Air Act. Benzene alone is also known to be carcinogenic.

A.3 | Monte Carlo simulation

Using a Monte Carlo simulation, we investigate whether our findings could result from spurious correlation between drilling activity and county-level disease trends. The simulation results indicate that the potential for unobservable county-level attributes to confound our findings in this manner is small.

In order to conduct the simulation, we first subset the sample to 2000–2005, the period before unconventional drilling in Marcellus shale began, and then randomly assign the "treatment" of unconventional wells to counties that contained wells in the post-2005 period. These wells are assigned by drawing the year of drilling in the prewell period from a uniform distribution. We then estimate Equation 1 for each condition with current and lagged placebo treatment indicators but without the output variables.

Provided there are no unobserved determinants of hospitalization rates in well counties, the coefficients on the placebo treatment variables should be statistically significant under a t test at the same rate at the α level of the test (i.e., the Type 1 error rate). If, however, we find a spurious correlation between the placebo treatment and hospitalization rates at a substantially higher rate than the α level of the test, it indicates that there are unobservable differences in factors determining hospitalization rates that are not accounted for by our models, which may lead us to incorrectly conclude that there is a statistically significant relationship between shale gas development and health. We report the 1%, 5%, and 10% rejection rates for the placebo treatments for the models corresponding to our health outcomes in Table 5.

A.4 | Synthetic control method

In this section, we provide some technical detail on how we implement the synthetic control method and conduct statistical inference. Following Abadie et al. (2010), suppose county j has the first unconventional well drilled in year T_0 (i.e., "treatment"), and we observe T years in total. Denote Y_{jt}^0as the outcome of interest in the absence of unconventional wells and Y_{jt}^1as the outcome when there are active unconventional wells in the county. Then, for each posttreatment year $t = T_0 + 1, \ldots, T$, the treatment effect can be written as $\alpha_{jt} = Y_{jt}^1 - Y_{jt}^0$. To estimate the counterfactuals Y_0^tin the posttreatment years, we construct a "synthetic" county j by selecting a set of nonnegative weights W for control counties ($g = 1, 2, \ldots, G$). To measure the discrepancy between the treatment county and its synthetic match, we use the following distance metric:

$$\sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)} \tag{A.1}$$

where X_1 is a vector that includes county-level predictors and pretreatment outcomes for the treatment county, and X_0 is the corresponding matrix for the control counties. The symmetric matrix V assigns weights to variables included in X. Ideally, the synthetic match should approximate the pretreatment outcome trajectory of the treatment county as closely as possible and at the same time have similar county characteristics. To achieve this, Abadie et al. (2010) suggest using a constrained optimization procedure that jointly chooses optimal W and V such that the mean squared prediction error of the outcome variable is minimized for the pretreatment period:

$$MSPE = (Y_1 - Y_0 W(V))' (Y_1 - Y_0 W(V)) \tag{A.2}$$

where Y_1 and Y_0 are vectors that contain the entire pretreatment outcome trajectory for the treatment and control counties. After obtaining the optimal weight W^*, the weighted averages of the observed outcome for the control counties is then used as the counterfactual in each of the posttreatment years. It follows that the estimator for the treatment effect in a given posttreatment year t can be written as:

TABLE A1 ICD-9-CM diagnosis codes for all five conditions

AMI
4100 41000 41001 41002 4101 41010 41011 41012 4102 41020 41021 41022 4103 41030 41031 41032 4104 41040 41041 41042 4105 41050 41051 41052 4106 41060 41061 41062 4107 41070 41071 41072 4108 41080 41081 41082 4109 41090 41091 41092
COPD
490 4910 4911 4912 49120 49121 49122 4918 4919 4920 4928 494 4940 4941 496
Asthma
49300 49301 49302 49310 49311 49312 49320 49321 49322 49381 49382 49390 49391 49392
Pneumonia
00322 0203 0204 0205 0212 0221 0310 0391 0521 0551 0730 0830 1124 1140 1144 1145 11505 11515 11595 1304 1363 4800 4801 4802 4803 4808 4809 481 4820 4821 4822 4823 48230 48231 48232 48239 4824 48240 48241 48242 48249 4828 48281 48282 48283 48284 48289 4829 483 4830 4831 4838 4841 4843 4845 4846 4847 4848 485 486 5130 5171
URI
4660 4661 46611 46619 0320 0321 0322 0323 0340 460 4610 4611 4612 4613 4618 4619 462 4640 46400 46401 46410 46411 46420 46421 46430 46431 4644 46450 46451 4650 4658 4659 4730 4731 4732 4733 4738 4739 78491

$$\widehat{\alpha}_{jt} = Y^1_{jt} - \sum_{g=1}^{g} w_g Y^0_{gt} \tag{A.3}$$

In our application, there are multiple counties that received the treatment in different years. We follow the approach by Cavallo et al. (2013) to obtain a single average treatment effect for each posttreatment year. Specifically, for a total of J well counties, the average effect in the K^{th} year after the treatment (i.e., K^{th} lead) is given by $\overline{\alpha}_k = \frac{1}{J}\sum_j \widehat{\alpha}_{jk}$.

We use the procedure described in Cavallo et al. (2013) to conduct statistical inference for $\overline{\alpha}_k$. The idea is to see whether $\overline{\alpha}_k$ is extreme when compared against an empirical distribution of *average* placebo effects. We begin by obtaining all placebo effects available to a given well county j.[24] This is done by iteratively assigning placebo treatment (and using the same treatment year as county j) to control counties and applying the synthetic control method outlined above using the remaining control counties to construct a synthetic match. After this step, we have J sets of placebo effects, each corresponding to a well county. Note that because we include in the control groups only those counties that do not have any unconventional wells by the end of 2013, all the well counties effectively have the same set of control counties. However, they still have different sets of placebo effects if they received treatment in different years.

Next, we construct average placebo effects from these J sets of G placebos. In total, there are $\prod_{j=1}^{J} G = G^J$ possible ways to obtain an average from the J well counties for a given lead K. The p value for a two-sided test of no effect can be calculated as:

$$p = \frac{\sum_{m=1}^{G^J} 1(|\overline{\alpha}_m| > |\overline{\alpha}_k|)}{G^J} \tag{A.4}$$

It is usually impractical to obtain all possible averages even when G or J is modestly large. In our application, we would have to obtain 28^8 = 3.778e + 11 average placebo effects. Therefore, we use 1,000 placebo averages to calculate p values, which we believe should be sufficiently large to uncover truly extreme effects.

[24]Given the number of placebo tests, we did not use fully constrained optimization to obtain the V matrix in our application. Instead, we used a regression-based V and obtained optimal W by minimizing (2), treating V as given. This significantly saves computational time at the cost of slightly worse fit for certain placebos.

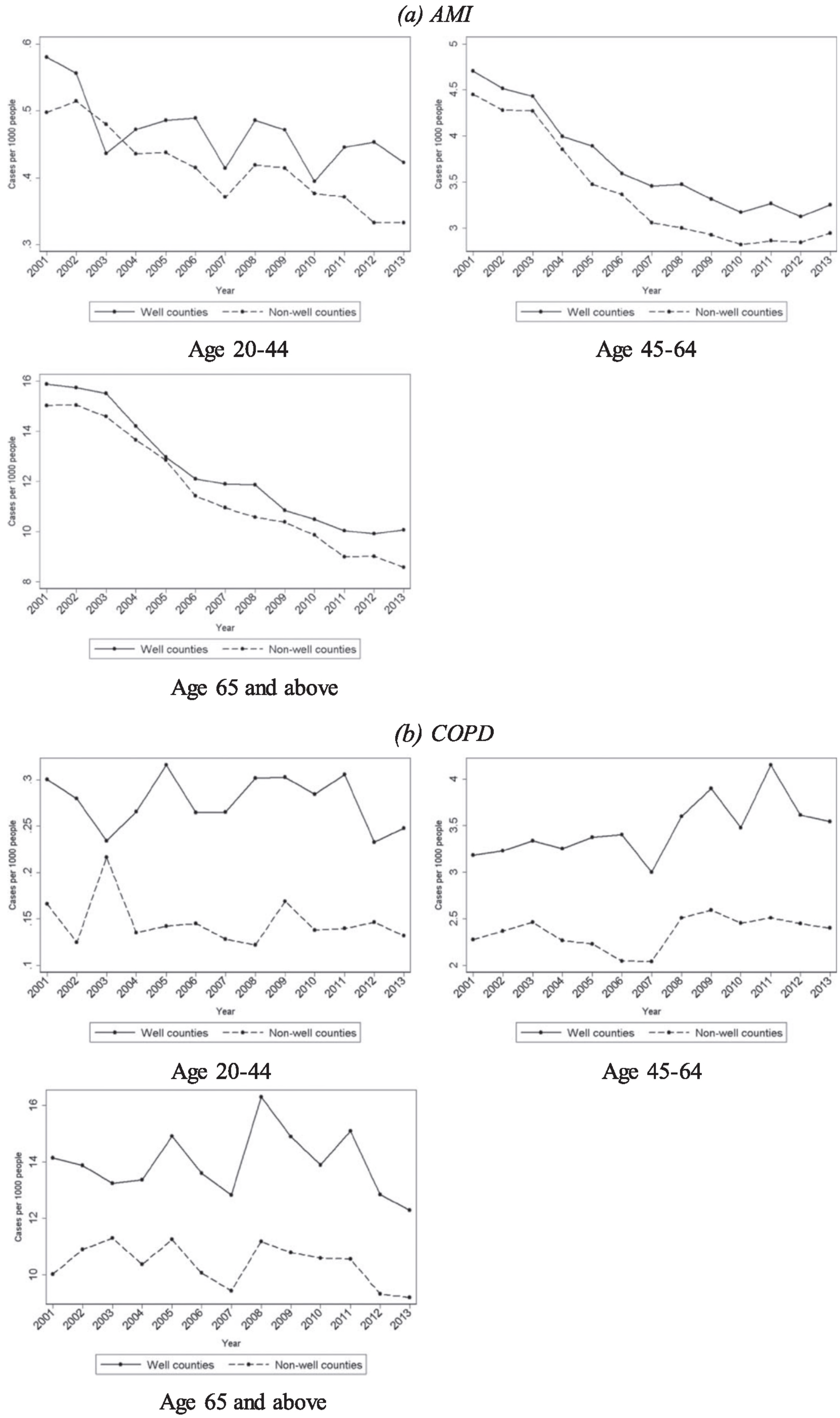
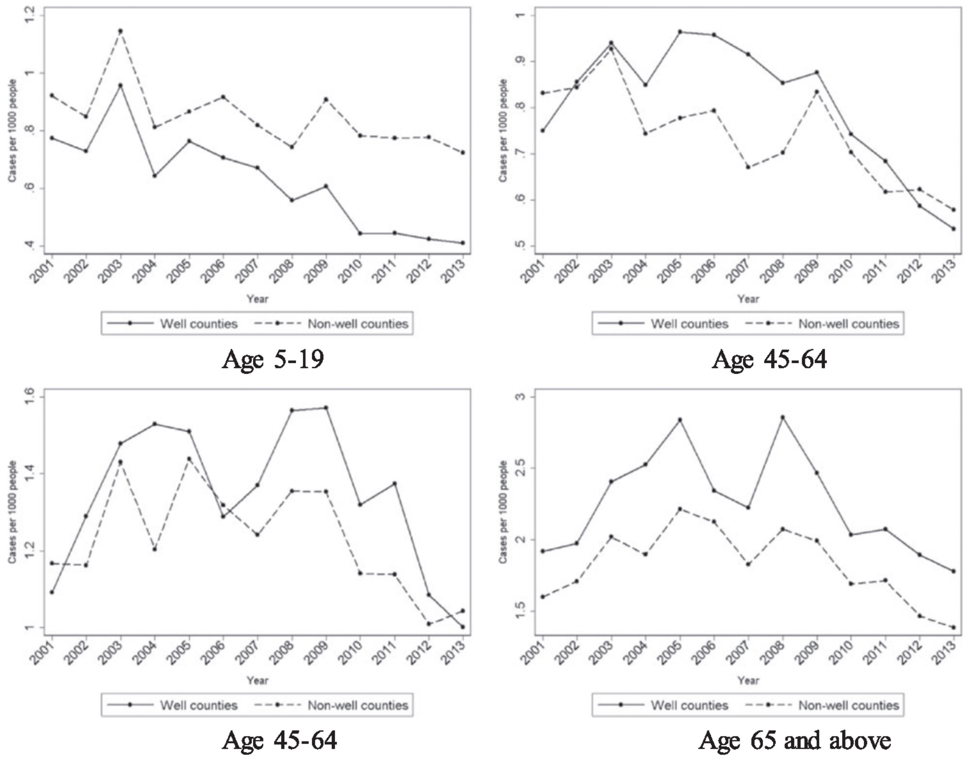


FIGURE A1 Hospitalization rates for select conditions, 2001–2013 *(c) Asthma*



(d) Pneumonia

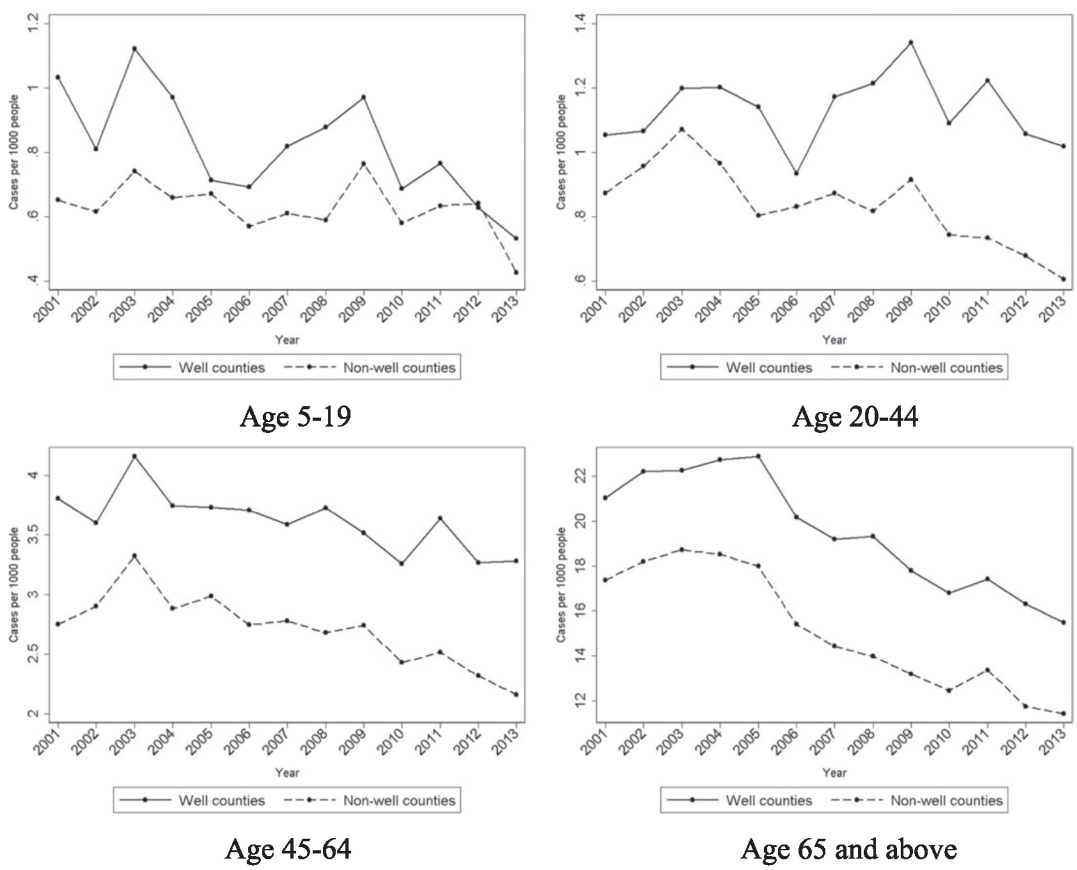


FIGURE A1 Continued.

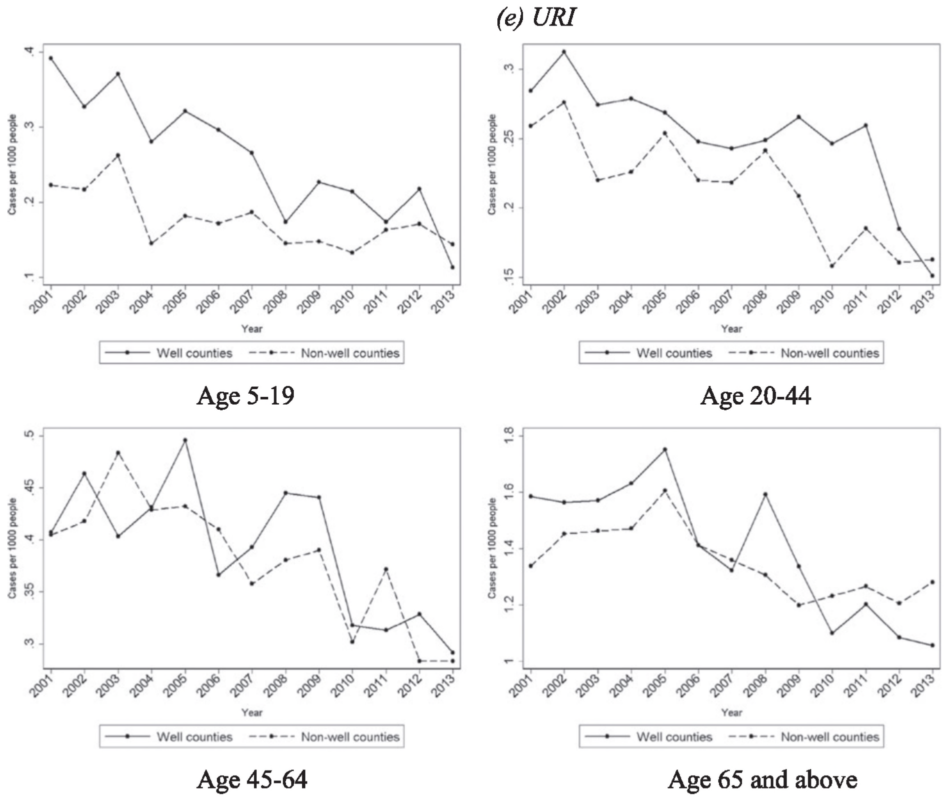


FIGURE A1 Continued.

TABLE A2 Impact of shale gas development on county-level hospitalization rates for other conditions, 2001–2013

	AMI		COPD		Asthma		URI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Full sample (age 5 and above)								
Well current year	−0.027 (0.074)	−0.072 (0.083)	0.123 (0.098)	−0.018 (0.104)	0.004 (0.053)	−0.038 (0.051)	0.051 (0.033)	0.035 (0.030)
Well last year	0.084 (0.079)	0.006 (0.075)	−0.091 (0.115)	−0.091 (0.111)	0.051 (0.038)	−0.008 (0.045)	−0.015 (0.039)	−0.025 (0.044)
Log output	0.002 (0.009)	0.007 (0.009)	−0.001 (0.011)	−0.002 (0.008)	−0.003 (0.005)	0.004 (0.006)	−0.003 (0.003)	0.000 (0.004)
1st lag of log output	0.019** (0.008)	0.011 (0.010)	0.007 (0.009)	−0.006 (0.013)	0.000 (0.004)	−0.003 (0.004)	0.001 (0.003)	−0.001 (0.003)
	−0.027	-0.072	0.123	−0.018	0.004	−0.038	0.051	0.035
Panel B. Age 5–19								
Well current year					−0.000 (0.064)	0.000 (0.069)	−0.044 (0.041)	−0.014 (0.042)
Well last year					−0.052 (0.061)	−0.084 (0.073)	0.011 (0.041)	0.034 (0.050)
Log output					0.000 (0.006)	0.003 (0.007)	0.001 (0.005)	0.001 (0.006)
1st lag of log output					−0.005 (0.006)	−0.005 (0.006)	−0.001 (0.003)	−0.003 (0.004)
Panel C. Age 20–44								
Well current year	−0.058 (0.043)	−0.065 (0.050)	0.057** (0.027)	0.074** (0.029)	−0.034 (0.046)	−0.051 (0.048)	0.062 (0.052)	0.072 (0.052)
Well last year	0.102** (0.049)	0.113* (0.059)	−0.078* (0.041)	−0.048 (0.045)	0.099** (0.049)	0.146*** (0.053)	−0.084 (0.054)	−0.070 (0.055)
Log output	−0.003 (0.004)	−0.003 (0.006)	0.000 (0.003)	−0.002 (0.003)	−0.010 (0.006)	−0.015** (0.006)	0.001 (0.003)	0.003 (0.004)
1st lag of log output	−0.000 (0.003)	−0.001 (0.004)	−0.001 (0.003)	−0.001 (0.003)	0.006 (0.005)	0.001 (0.006)	0.002 (0.002)	0.002 (0.003)
Panel D. Age 45–64								
Well current year	−0.076 (0.121)	−0.053 (0.133)	−0.009 (0.146)	−0.166 (0.146)	−0.001 (0.077)	−0.053 (0.081)	0.064 (0.039)	0.026 (0.039)
Well last year	−0.013 (0.141)	−0.032 (0.149)	−0.300* (0.170)	−0.227 (0.170)	0.047 (0.078)	−0.009 (0.078)	−0.041 (0.044)	−0.027 (0.047)
Log output	0.001 (0.013)	0.003 (0.015)	0.021 (0.015)	0.013 (0.014)	0.008 (0.007)	0.019*** (0.007)	−0.003 (0.004)	−0.004 (0.005)
1st lag of log output	0.019** (0.009)	0.011 (0.010)	−0.002 (0.016)	−0.019 (0.020)	−0.006 (0.005)	−0.013** (0.006)	0.002 (0.004)	−0.004 (0.004)
Panel E. Age 65 and above								
Well current year	0.061 (0.352)	−0.093 (0.352)	0.399 (0.375)	−0.119 (0.374)	−0.010 (0.163)	−0.092 (0.174)	0.101 (0.111)	0.011 (0.097)
Well last year	0.275 (0.390)	−0.020 (0.345)	0.369 (0.449)	0.043 (0.532)	0.106 (0.122)	−0.096 (0.129)	0.125 (0.104)	0.041 (0.117)
Log output	0.014 (0.048)	0.031 (0.045)	−0.030 (0.045)	−0.014 (0.037)	−0.011 (0.015)	0.007 (0.016)	−0.021* (0.010)	−0.003 (0.009)
1st lag of log output	0.070* (0.041)	0.052 (0.043)	0.013 (0.034)	−0.008 (0.046)	−0.002 (0.013)	0.006 (0.012)	0.001 (0.009)	0.003 (0.010)
Year fixed effects	X	X	X	X	X	X	X	X
County specific linear trends		X		X		X		X

Note. Standard errors are clustered at the county level. The total of number of observations is 804. All models include county and year fixed effects. Other control variables include average age, the share of different types of insurance (Medicare, Medicaid, private, self-pay, government, and other insurance), the share of female patients, the share of different race and ethnicity groups (White, Black, Asian, Hispanic, and other race), the share of different types of admission (emergency, urgent, elective, and other types), average Charlson index, county-level unemployment rate, poverty rate, annual quartiles of median household income, log of population density, log of annual coal production (both surface and underground), log of number of conventional wells, log of conventional output and the entire county-level age distribution.

$^*p < .1$.$^{**}p < .05$.$^{***}p < .01$.

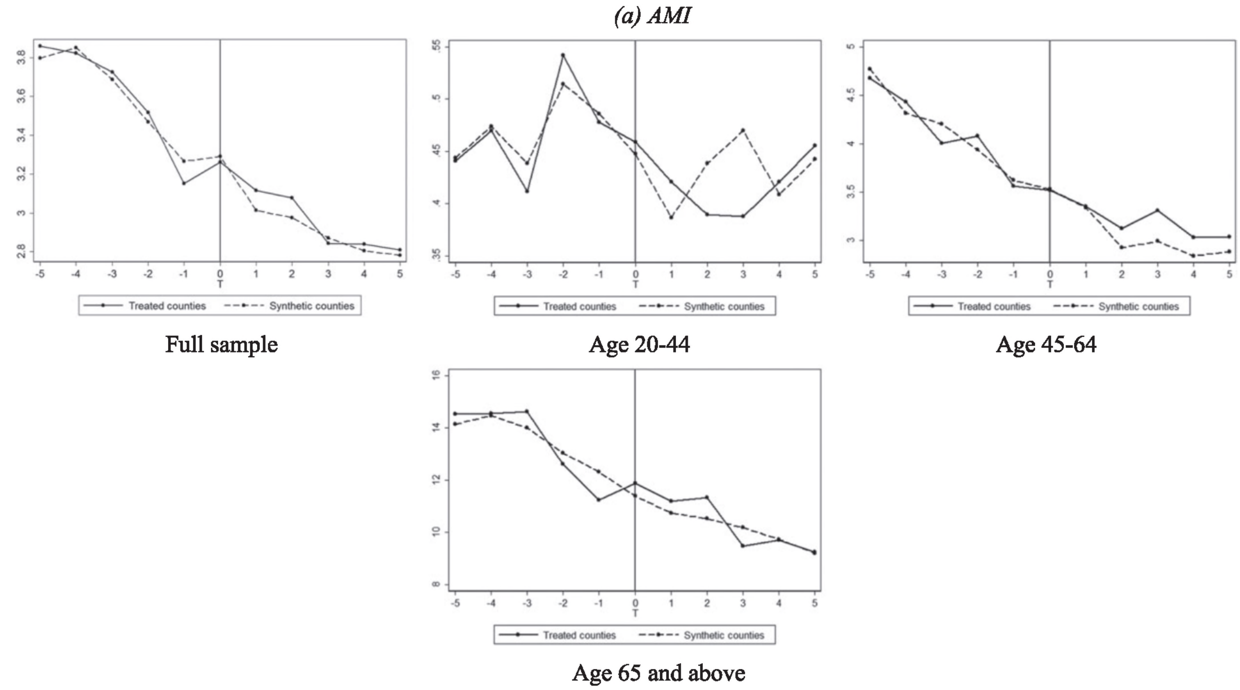


FIGURE A2 Evolution of hospitalization rates in well counties in the top quartile of the well distribution and their synthetic matches

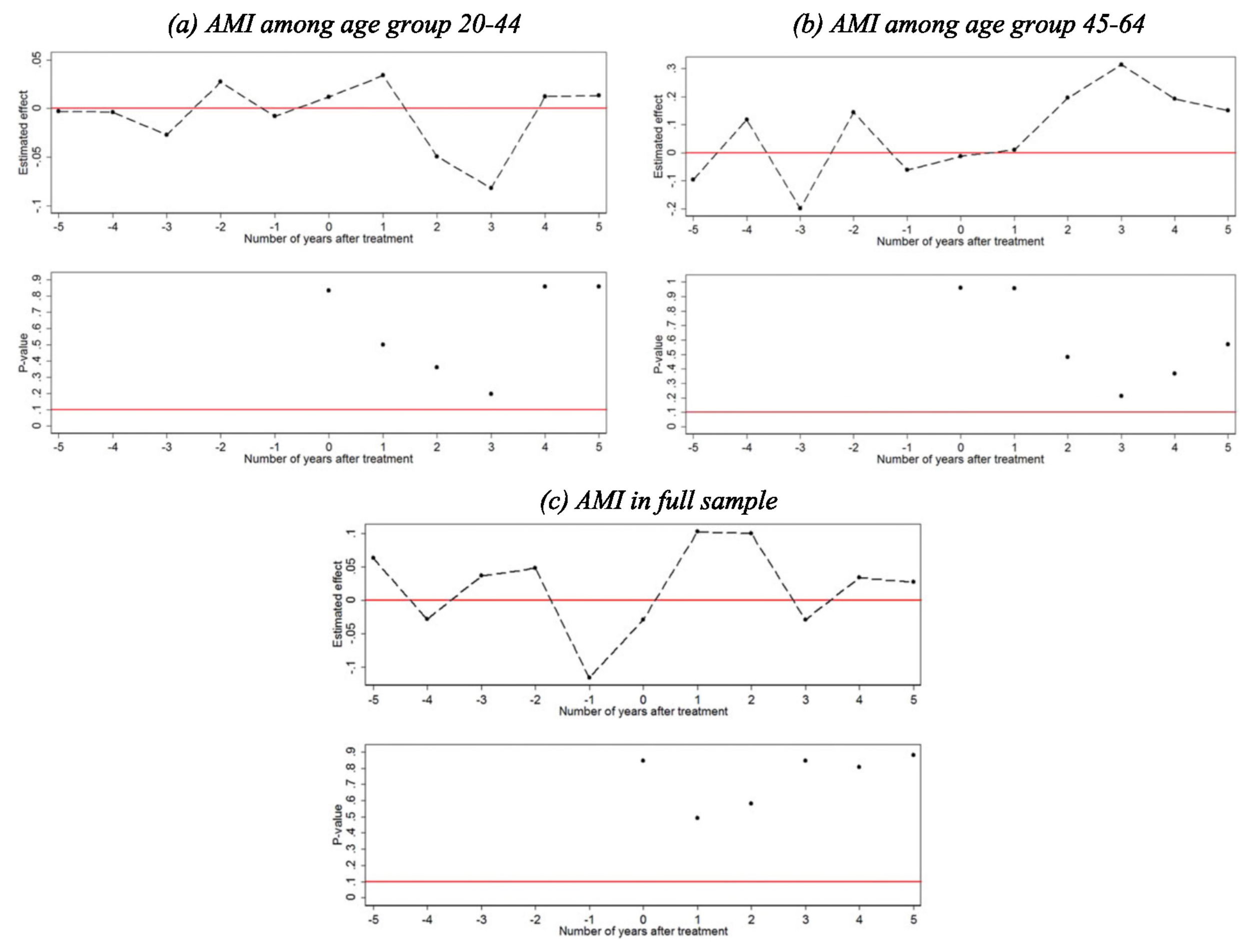


FIGURE A3 *P* values of synthetic control estimates for select outcomes and age groups [Colour figure can be viewed at wileyonlinelibrary.com]

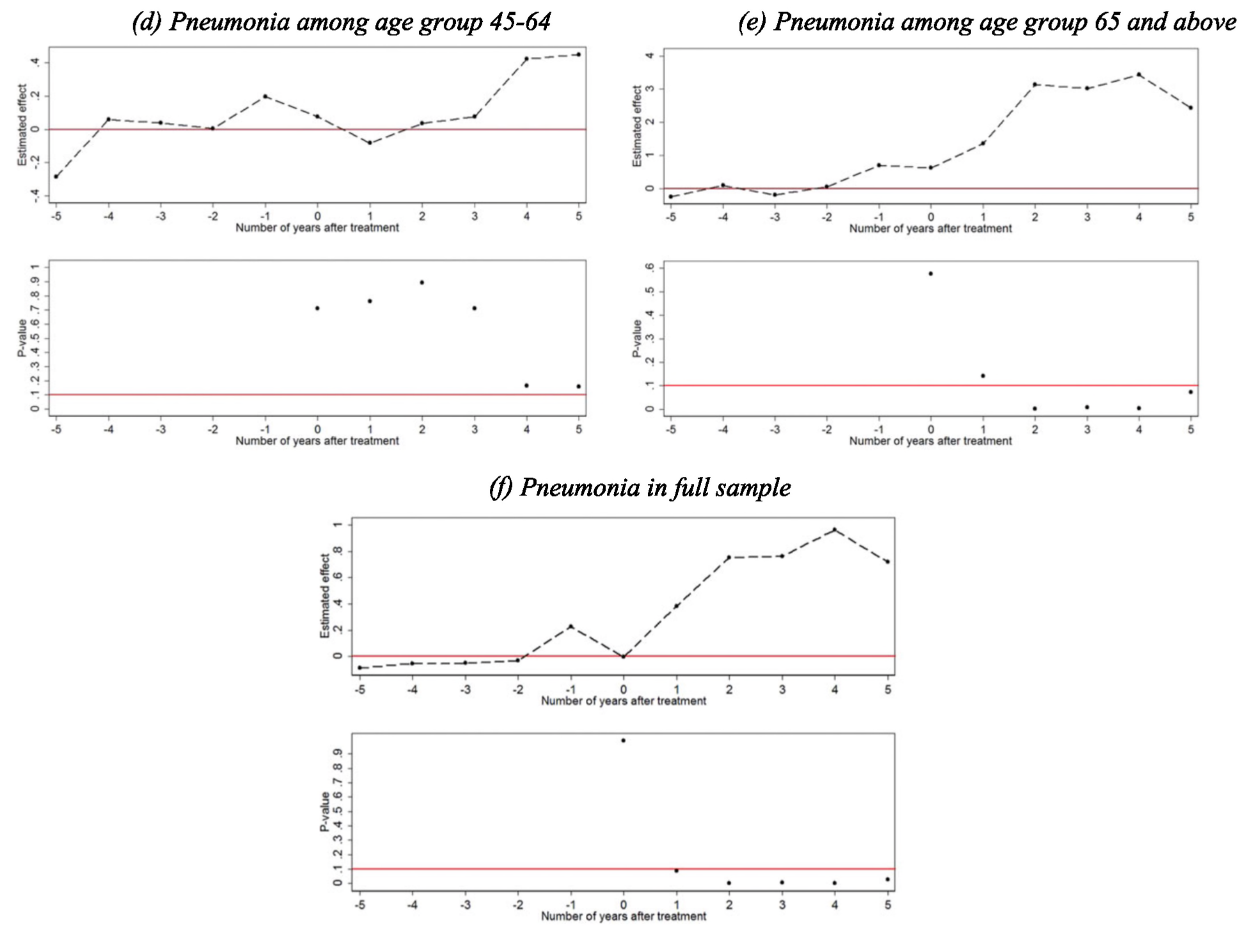


FIGURE A3 Continued.

TABLE A3 Coefficient estimates from pretrend tests

	(1) AMI	(2) COPD	(3) Asthma	(4) Pneumonia	(5) URI
Panel A. Full sample (Age 5 and above)					
Well current	−0.033 (0.082)	0.052 (0.120)	0.030 (0.051)	0.282* (0.155)	0.052* (0.028)
1st lead of well current	0.083 (0.106)	−0.042 (0.138)	0.027 (0.041)	0.030 (0.142)	−0.027 (0.034)
2nd lead of well current	−0.054 (0.073)	0.240* (0.130)	0.006 (0.052)	0.142 (0.142)	0.028 (0.028)
Panel B. Age 5–19					
Well current			−0.026 (0.050)	0.008 (0.064)	0.026 (0.034)
1st lead of well current			−0.020 (0.075)	0.005 (0.075)	−0.071* (0.039)
2nd lead of well current			−0.039 (0.067)	0.070 (0.074)	0.014 (0.036)
Panel C. Age 20–44					
Well current	−0.005 (0.049)	0.039 (0.029)	0.029 (0.053)	0.012 (0.072)	0.025 (0.030)
1st lead of well current	0.007 (0.051)	−0.042 (0.040)	0.004 (0.064)	0.112 (0.094)	0.008 (0.027)
2nd lead of well current	−0.013 (0.038)	−0.003 (0.036)	0.033 (0.053)	0.011 (0.082)	−0.012 (0.028)
Panel D. Age 45–64					
Well current	0.012 (0.138)	−0.065 (0.170)	0.047 (0.089)	0.051 (0.167)	0.007 (0.037)
1st lead of well current	−0.088 (0.179)	−0.221 (0.235)	0.087 (0.089)	0.034 (0.184)	0.023 (0.045)
2nd lead of well current	−0.029 (0.158)	0.368* (0.205)	−0.055 (0.071)	0.167 (0.147)	0.029 (0.037)
Panel E. Age 65 and above					
Well current	−0.019 (0.351)	0.309 (0.451)	−0.019 (0.150)	1.313** (0.615)	0.148 (0.111)
1st lead of well current	0.337 (0.405)	0.233 (0.506)	−0.012 (0.121)	0.002 (0.572)	−0.148 (0.141)
2nd lead of well current	−0.284 (0.338)	0.419 (0.506)	0.060 (0.157)	0.285 (0.558)	0.109 (0.127)

Note. Standard errors are clustered at the county level. The total of number of observations is 737. All models include county and year fixed effects. Other control variables include average age, the share of different types of insurance (Medicare, Medicaid, private, self-pay, government, and other insurance), the share of female patients, the share of different race and ethnicity groups (White, Black, Asian, Hispanic, and other race), the share of different types of admission (emergency, urgent, elective, and other types), average Charlson index, county-level unemployment rate, poverty rate, annual quartiles of median household income, log of population density, log of annual coal production (both surface and underground) and the entire county-level age distribution.

$^{*}p < .1$. $^{**}p < .05$. $^{***}p < .01$.

TABLE A4 Synthetic control estimates for 8 counties in the top quartile of the unconventional well distribution

	$T = 0$	$T = 1$	$T = 2$	$T = 3$	$T = 4$	$T = 5$	RMSPE	Mean of outcome	RMSPE/Mean (%)
Panel A. Full sample (Age 5 and above)									
AMI	−0.029	0.103	0.100	−0.029	0.034	0.027	0.191	3.307	**6**
COPD	0.380	0.618	1.093	1.236	1.093	1.368	0.337	3.157	11
Asthma	0.053	0.152	0.136	−0.038	0.062	0.220	0.130	1.144	11
Pneumonia	−0.003	0.383	0.754	0.763	0.964	0.719	0.175	4.544	**4**
URI	0.001	0.036	0.000	0.062	0.001	0.034	0.057	0.478	12
Panel B. Age 5–19									
Asthma	−0.130	−0.106	−0.086	−0.296	−0.205	−0.367	0.153	0.719	21
Pneumonia	−0.125	0.098	0.018	−0.165	0.028	−0.110	0.150	0.737	20
URI	0.140	0.142	−0.115	0.005	−0.016	0.033	0.097	0.225	43
Panel B. Age 20–44									
AMI	0.012	0.034	−0.049	−0.082	0.012	0.013	0.033	0.447	***7***
COPD	0.036	0.160	0.196	0.075	0.187	0.097	0.115	0.223	52
Asthma	0.070	0.232	0.096	−0.064	−0.052	0.090	0.149	0.780	19
Pneumonia	0.009	0.074	0.247	0.092	0.271	0.383	0.149	1.008	15
URI	0.086	0.034	−0.030	0.070	0.069	0.100	0.089	0.236	38
Panel C. Age 45–64									
AMI	−0.012	0.011	0.195	0.315	0.192	0.152	0.300	3.577	***8***
COPD	0.587	0.394	0.490	0.804	0.870	0.760	0.324	3.001	11
Asthma	0.046	0.283	0.059	0.128	0.190	0.336	0.187	1.296	14
Pneumonia	0.077	−0.083	0.036	0.077	0.424	0.452	0.245	3.237	***8***
URI	0.037	0.065	0.060	0.079	0.143	0.017	0.112	0.387	29

(Continues)

TABLE A4 (Continued)

	$T = 0$	T = 1	$T = 2$	$T = 3$	$T = 4$	$T = 5$	RMSPE	Mean of outcome	RMSPE/Mean (%)
Panel D. Age 65 and above									
AMI	0.482	0.455	0.815	−0.707	−0.031	0.044	1.225	12.080	10
COPD	0.948	2.107	3.932	4.168	3.403	4.608	1.451	12.450	12
Asthma	0.033	0.206	0.609	−0.207	0.088	0.500	0.382	2.075	18
Pneumonia	0.625	1.356	3.138	3.013	3.428	2.434	0.810	17.686	***5***
URI	−0.091	0.017	0.291	0.194	−0.329	−0.093	0.273	1.380	20

Note. The eight counties included in this table are Bradford, Butler, Fayette, Greene, Lycoming, Susquehanna, Tioga, and Westmoreland. T denotes the relative time to "treatment" (i.e., the year when the first unconventional well was drilled). The estimate in each cell represents the difference between the average outcome in the treatment counties and the average outcome in the synthetic matches. RMSPE represents the average root mean squared prediction error, which measures the pretreatment fit between the treatment counties and their respective synthetic matches. The bold and italic entries are used to indicate values less than 10.

TABLE A5 Rejection rates for H_0 from Monte Carlo simulations of placebo treatments, 2000–2005

	Full sample			Age 5–19			Age 20–44			Age 45–64			Age 65 and above		
	<0.01	<0.05	<0.1	<0.01	<0.05	<0.1	<0.01	<0.05	<0.1	<0.01	<0.05	<0.1	<0.01	<0.05	<0.1
AMI															
Well current year (%)	1.56	7.10	13.30				1.04	5.85	11.61	1.56	6.63	12.25	1.65	7.32	13.75
Well last year (%)	1.90	7.29	13.50				1.04	5.81	11.57	1.44	6.39	12.32	1.62	7.80	13.97
COPD															
Well current year (%)	1.03	5.88	11.97				1.56	7.76	14.15	1.86	7.53	13.64	0.78	4.75	10.09
Well last year (%)	1.15	6.37	12.52				1.68	7.27	13.83	1.81	7.49	13.72	0.62	4.29	9.74
Asthma															
Well current year (%)	6.34	18.13	26.91	1.33	6.50	12.09	4.01	12.62	20.54	1.53	6.52	12.22	3.36	12.21	20.55
Well last year (%)	7.32	19.80	29.87	1.36	6.02	12.08	4.95	14.08	21.83	2.00	7.71	13.78	3.28	11.77	19.93
Pneumonia															
Well current year (%)	1.90	8.66	16.07	1.07	5.67	11.25	1.22	5.86	11.22	1.44	7.02	14.35	1.28	6.47	12.87
Well last year (%)	1.87	8.73	16.44	1.00	5.22	10.62	1.24	5.90	11.36	1.17	6.64	13.52	1.25	6.78	13.37
URI															
Well current (%)	0.93	5.54	10.64	1.35	6.66	12.31	1.10	5.39	10.54	0.98	5.00	10.00	2.17	8.61	14.90
Well last year (%)	0.95	4.84	9.98	1.19	6.37	12.20	0.98	5.14	10.37	0.97	4.72	9.68	1.47	7.19	14.14

Note. Results are based on 10,000 replications. All models include county and year fixed effects. Control variables include average age, the share of different types of insurance (Medicare, Medicaid, private, self-pay, government, and other insurance), the share of female patients, the share of different race and ethnicity groups (White, Black, Asian, Hispanic, and other race), the share of different types of admission (emergency, urgent, elective, and other types), average Charlson index, county-level unemployment rate, poverty rate, annual quartiles of median household income, log of population density, log of annual coal production (both surface and underground), log of number of conventional wells, log of conventional output, and the entire county-level age distribution.

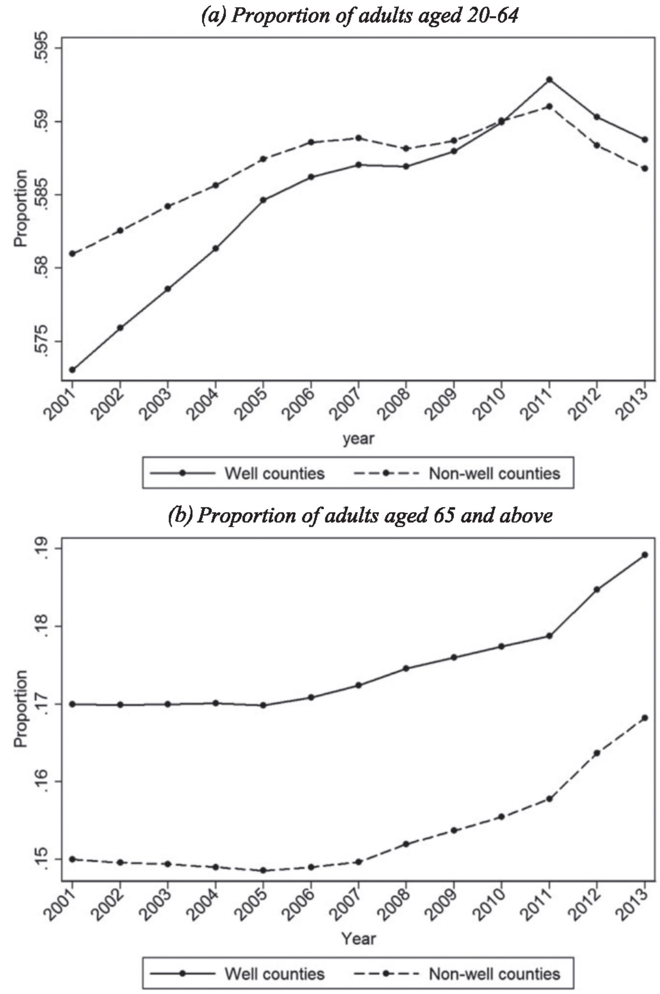


FIGURE A4 Changes in the county age distribution between 2001 and 2013