




Proximity to Oil Refineries and Risk of Cancer: A Population-Based Analysis

Stephen B. Williams , MD, MS,^{1,*} Yong Shan, PhD,¹ Usama Jazzar, MD,¹ Preston S. Kerr , MD,¹ Ikenna Okereke, MD,² V. Suzanne Klimberg, MD, PhD,³ Douglas S. Tyler, MD,⁴ Nagireddy Putluri, PhD,⁵ David S. Lopez , MS, MPH, DrPH,⁶ John D. Prochaska, DrPH, MPH,⁶ Cornelis Elferink, PhD,⁷ Jacques G. Baillargeon, PhD,⁸ Yong-Fang Kuo, PhD,⁸ Hemalkumar B. Mehta, PhD^{4,9}

¹Department of Surgery, Division of Urology, The University of Texas Medical Branch, Galveston, TX, USA; ²Department of Surgery, Division of Thoracic Surgery, The University of Texas Medical Branch, Galveston, TX, USA; ³Department of Surgery, Division of Surgical Oncology, The University of Texas Medical Branch at Galveston, Galveston, TX, USA; ⁴Department of Surgery, The University of Texas Medical Branch at Galveston, Galveston, TX, USA; ⁵Department of Molecular and Cellular Biology, Dan L. Duncan Cancer Center, Advanced Technology Core, Alkek Center for Molecular Discovery, Baylor College of Medicine, Houston, TX, USA; ⁶Department of Preventive Medicine and Population Health, The University of Texas Medical Branch at Galveston, Galveston, TX, USA; ⁷Department of Pharmacology and Toxicology, Center for Environmental Toxicology, The University of Texas Medical Branch at Galveston, Galveston, TX, USA; ⁸Department of Medicine, Division of Epidemiology, Sealy Center on Aging, The University of Texas Medical Branch at Galveston, Galveston, TX, USA and ⁹Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

*Correspondence to: Stephen B. Williams, MD, MS, Division of Urology, The University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555, USA (e-mail: stbwilli@utmb.edu).

Abstract

Background: The association between proximity to oil refineries and cancer rate is largely unknown. We sought to compare the rate of cancer (bladder, breast, colon, lung, lymphoma, and prostate) according to proximity to an oil refinery in Texas. **Methods:** A total of 6 302 265 persons aged 20 years or older resided within 30 miles of an oil refinery from 2010 to 2014. We used multilevel zero-inflated Poisson regression models to examine the association between proximity to an oil refinery and cancer rate. **Results:** We observed that proximity to an oil refinery was associated with a statistically significantly increased risk of incident cancer diagnosis across all cancer types. For example, persons residing within 0-10 (risk ratio [RR] = 1.13, 95% confidence interval [CI] = 1.07 to 1.19) and 11-20 (RR = 1.05, 95% CI = 1.00 to 1.11) miles were statistically significantly more likely to be diagnosed with lymphoma than individuals who lived within 21-30 miles of an oil refinery. We also observed differences in stage of cancer at diagnosis according to proximity to an oil refinery. Moreover, persons residing within 0-10 miles were more likely to be diagnosed with distant metastasis and/or systemic disease than people residing 21-30 miles from an oil refinery. The greatest risk of distant disease was observed in patients diagnosed with bladder cancer living within 0-10 vs 21-30 miles (RR = 1.30, 95% CI = 1.02 to 1.65), respectively. **Conclusions:** Proximity to an oil refinery was associated with an increased risk of multiple cancer types. We also observed statistically significantly increased risk of regional and distant/metastatic disease according to proximity to an oil refinery.

Expansion of oil production in the United States has resulted in a record-high production of 18.8 million barrels per day, with Texas as the largest producer (1). Debate over potential health risks, primarily ensuing from emitted air pollutants, has been increasingly underscored in the public and legislative arenas (2-4). The International Agency for Research on Cancer has recognized and identified important sources of emissions from the petrochemical industries as environmental carcinogens (International Agency for Research on Cancer: monographs on the evaluation of carcinogenic risks to humans) (5).

Prior studies have been conducted to investigate the rate of different cancers among residents who live near oil refineries (ie, petrochemical industrial complexes [PIC]). Lung cancer has been previously linked with polycyclic aromatic hydrocarbons exposure (6), with conflicting results on the risks of lung and other cancer sites according to PIC proximity (7-13). However, most studies have been conducted outside of the United States—in the setting of different sociodemographic characteristics for at-risk residents—which in turn make results difficult to extrapolate to the United States (14). Given this impetus, we

Received: 14 April 2020; Revised: 11 August 2020; Accepted: 22 September 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

sought to test the hypothesis that there is an increased cancer risk across different cancer types according to proximity to an oil refinery in Texas.

Methods

Data Sources

To calculate cancer rate, we used 2 primary data sources. First, cancer case patients from 2001 to 2014 were identified using International Classification of Diseases for Oncology–3 codes from the Texas Cancer Registry (TCR), a statewide population-based registry that meets the National Program of Central Cancer Registries and Centers for Disease Control and Prevention high-quality data standards (15,16). Second, we used the US Census Bureau database American Community Survey (ACS) 5-year estimates from 2010 to 2014 to derive the denominator population (17). We selected ACS 2010–2014 because it provided stable population estimates for 5 years and captured important variables on sociodemographic characteristics. We used the County Health Rankings and Roadmaps database to capture air pollution and smoking rates. The earliest available air pollution data were from 2008, and the earliest available smoking rates were from 2014. Therefore, we used different years of County Health Rankings and Roadmaps databases. Smoking rates were reported in years prior to 2014, but more than two-thirds of counties had missing values; 2014 was the first year with complete data. We used the Railroad Commission of Texas database from 2003 to capture the number of oil wells. [Supplementary Table 1](#) (available online) provides a description of all data sources (18). The health rankings model was established in 2010 to discern factors contributing to differences in overall health outcomes across various US counties. Although no individually employed model may perfectly describe different existing populations, the ranking model has been shown to provide relatively accurate information on factors responsible for worsening health outcomes and ultimately helps raise awareness for public health improvement (19–21). This study was approved by the institutional review board at the University of Texas Medical Branch.

Exposure

The primary exposure of interest was distance from an oil refinery to residence zip code. A total of 28 active oil refineries were identified from 2001 to 2014 in Texas, and the respective zip code for each address was then collected ([Supplementary Table 2](#), available online). We used ArcGIS software (ESRI, Redlands, CA), a proprietary geographical information systems (GIS) tool, to compute the linear distance (also known by the phrase “as the crow flies”) from the center of the residence zip code to the center of the zip code of the nearest oil refinery (22,23). For example, all people living in the same zip code as the refinery will have a distance of 0 miles. All persons residing outside a 30-mile radius from an oil refinery were excluded to limit further potential confounding from exposure to industry pollutants from textile, chemical, or fertilizer facilities (24,25).

Outcomes

Our primary outcome was cancer rates. This was calculated for all adults (20 years or older) with a new diagnosis of cancer of the bladder, prostate, breast, lung, colon, or lymphoma from

January 1, 2001, through December 31, 2014. We measured cancer rate by proximity to an oil refinery. Each cancer rate was stratified based on the distance from an oil refinery. Our selection of cancers was based on the highest US rates and malignancies included in previous oil refinery proximity studies. Prostate and breast cancers continue to be the malignancies with highest rates in men and women, respectively (26). As aforementioned, in studies investigating oil refinery proximity to cancer rates, efforts were largely focused on lung and lymphohematopoietic malignancies. Colon and bladder cancer rates follow in ranking per the most recent US data and were included in the earliest US efforts to elucidate pollutant harm in PIC regions (26,27).

We used diagnosis of fracture (International Classification of Diseases for Oncology–9 codes: 800–829) as a negative control outcome. These diagnoses were selected because this medical condition was deemed unlikely to be related to proximity to an oil refinery. For the control outcomes, we used the 100% Texas Medicare database and selected participants aged 65 years or older with any incident fracture diagnosis from 2010 to 2014 and who resided within 30 miles of an oil refinery. The denominator was restricted to the ACS population, which resided within 30 miles of an oil refinery.

Study Covariates

From TCR, we extracted patient clinical information. Data collected at the zip code level included patient demographic characteristics, primary tumor type, primary tumor site, stage, and tumor morphology. For each zip code, we used the following variables: age (20–39, 40–59, 60 years or older), sex (male, female), race and ethnicity (White, Black, others), median household income (categorized into quartiles), and education (high school diploma or higher education; categorized into quartiles); all variables were obtained from the 2010 to 2014 ACS. The County Health Rankings and Roadmaps database was used to capture each county’s level of current smoking rates and atmospheric particulate matter (PM_{2.5}) level (18). High averages of fine particulates (particles 2.5 micrometers or less in size in micrograms per cubic meter) from the air is a proxy for air pollution, and the PM_{2.5} level is a validated measure for air pollution taken from 2008 (28). Because of the relatively large number of oil wells in Texas, we also identified all oil wells at the county level using the Railroad Commission of Texas database from 2003 (29).

Statistical Analysis

Population-level characteristics, PM_{2.5}, and number of oil wells were, for the calendar years 2010–2014, reported by proximity to oil refinery. We calculated age-standardized cancer rates per 10 000 persons for 2001–2014 and control (fracture) outcome rates for 2010–2014; to calculate rates, the numerator included all patients with the outcome, and the denominator included ACS population from 2010 to 2014. After calculating age and group-specific rates of cancer, we weighted the rates by age group in the entire study cohort to calculate age-standardized rates. All analyses were conducted at zip code level. Several zip codes had no case patients of cancers. All covariates were measured at the county level except PM_{2.5}, smoking rates, and number of oil wells, which were measured at county level. To account for the nesting of zip codes within county, we used zero-inflated multilevel Poisson regression models to determine the association of proximity to oil refinery (0–10, 11–20, or 21–30

Table 1. Characteristics of census-derived population based on distance from an oil refinery, 2010-2014

Characteristics	Total population No. (%)	Distance from oil refineries ^a		
		0-10 miles No. (%)	11-20 miles No. (%)	21-30 miles No. (%)
Total population	6 302 265 (100)	2 213 854 (100)	2 433 988 (100)	1 654 423 (100)
Age group, y				
20-39	2 587 041 (41.0)	923 875 (41.7)	1 022 437 (42.0)	640 729 (38.7)
40-59	2 341 663 (37.2)	793 942 (35.9)	890 541 (36.6)	657 180 (39.7)
≥60	1 373 329 (21.8)	496 037 (22.4)	520 999 (21.4)	356 293 (21.5)
Sex				
Male	3 087 292 (49.0)	1 089 118 (49.2)	1 188 656 (48.8)	809 518 (48.9)
Female	3 214 740 (51.0)	1 124 736 (50.8)	1 245 321 (51.2)	844 683 (51.1)
Race and ethnicity				
White	4 388 769 (69.6)	1 512 062 (68.3)	1 698 916 (69.8)	1 177 791 (71.2)
Black	931 189 (14.8)	376 355 (17.0)	348 059 (14.3)	206 775 (12.5)
Other	982 074 (15.6)	325 437 (14.7)	387 002 (15.9)	269 635 (16.3)
Median household income (quartile)				
\$0-\$37 506	1 139 376 (18.1)	669 563 (30.2)	445 355 (18.3)	24 458 (1.5)
\$37 507-\$46 911	1 508 694 (23.9)	662 308 (29.9)	554 752 (22.8)	291 634 (17.6)
\$46 912-\$60 026	1 210 634 (19.2)	324 693 (14.7)	603 323 (24.8)	282 618 (17.1)
>\$60 026	2 443 330 (38.8)	557 291 (25.2)	830 548 (34.1)	1 055 491 (63.8)
Smoke	6 302 265 (100)	323 223 (14.6)	338 324 (13.9)	233 274 (14.1)
High school diploma or higher education (quartile)				
0-79.7	2 384 283 (37.8)	1 152 421 (52.0)	849 689 (34.9)	382 174 (23.1)
79.8-87.3	1 271 417 (20.2)	498 682 (22.5)	344 686 (14.2)	428 049 (25.9)
87.4-92.5	1 175 725 (18.7)	262 468 (11.9)	691 563 (28.4)	221 694 (13.4)
>92.5	1 470 607 (23.3)	300 283 (13.6)	548 039 (22.5)	622 285 (37.6)
Pollution PM2.5, mean (SD) ^b	6 302 265 (100)	9.4 (0.3)	9.3 (0.4)	9.3 (0.5)
No. of oil wells, mean (SD)	6 302 265 (100)	1570.0 (1421.8)	1933.0 (1518.2)	1424.6 (1295.9)

^aWe used ArcGIS software (ESRI, Redlands, CA) to compute the linear distance (also known as “as the crow flies”) from the center of the residence zip code of each study subject to the center of the zip code of the nearest oil refinery identified during the study period.

^bAtmospheric particulate matter (PM2.5) level measures high averages of fine particulates (particles 2.5 micrometers or less in size in micrograms per cubic meter) from the air taken as an average level in each county from 1998 when levels were first recorded to 2017.

miles) with cancer risk while controlling for covariates. We also accounted for spatial autocorrelation in these models by applying an unstructured covariance matrix; 6 regression models were constructed for cancers and 1 model for control outcome. In the first Poisson regression model, we modeled bladder cancer rates, and we constructed another 5 models for each cancer (prostate, breast, lung, colon, or lymphoma). The seventh model included fracture rates as an outcome. All Poisson regression models used total population size as offset and estimated rate ratios and corresponding 95% confidence intervals (CIs) (30,31). All models controlled for zip code level age, sex, percent of White race household income (socioeconomic status), county-level smoking rate, PM2.5 (air pollution), and total number of oil wells. Sex was not included for regression models for breast cancer (restricted to females) and prostate cancer (restricted to males). Age, sex, household income, and education were modeled as categorical variables, and White, smoking rates, PM2.5, and number of oil wells as continuous variables. We assessed linear trends in proximity to oil refinery with cancer risk using the linear contrast from the multilevel Poisson regression model. In the secondary analysis, we evaluated the association of proximity to an oil refinery with cancer stage at diagnosis as a measure of disease burden.

We conducted 3 sensitivity analyses. First, to ensure cancer case patients correspond with the denominator ACS 2010-2014 population, we conducted a sensitivity analysis by selecting cancer case patients from 2010 to 2014. Second, we performed stratified analysis to case patients aged 60 years or older and 20

to 59 years because many of these cancers occur predominantly in older adults. Third, we performed stratified analysis in males and females to determine if there is heterogeneity of effect. To assess the potential impact of an unmeasured confounder on the observed association with cancer, fractures were used as control outcomes. Because control outcomes were determined from Medicare beneficiaries, we compared these rates with a subset of older patients (65 years or older) from the TCR cohort. All statistical tests were 2-sided, using SAS version 9.4 software (SAS Institute, Cary, NC) and ArcGIS. A *P* value of less than .05 was considered statistically significant. Also, if the 95% confidence interval did not include 1.00, we considered the result statistically significant when a *P* value was not calculated.

Results

The characteristics of census-derived population percentage based on distance from an oil refinery are summarized in Table 1. A total of 6 302 265 persons aged 20 years or older resided within 30 miles of an oil refinery from 2010 to 2014. The populations, according to 0-10, 11-20, and 21-30 miles from an oil refinery, were similar according to age, sex, race and ethnicity, smoking status, and PM2.5. There were differences in median household income with more individuals residing within 0-10 vs 20-30 miles of an oil refinery having an income below the poverty level (669 563 [30.2%] vs 24 458 [1.5%] persons), respectively. We also observed an increased number of oil wells

Table 2. Age-standardized rates of cancer by proximity to oil refinery

Distance from oil refinery, miles	Total population	No. of case patients (rate per 10 000 individuals)						
		Cancer patients						Control patients Fracture
		Bladder	Lung	Prostate ^a	Breast ^b	Colon	Lymphoma	
0-10	2 213 854	5537 (25.0)	23 382 (105.6)	20 876 (191.7)	28 962 (257.5)	13 036 (58.9)	15 305 (69.1)	63 077 (284.9)
11-20	2 433 988	5615 (23.1)	20 904 (85.9)	21 921 (184.4)	31 265 (251.1)	11 677 (48.0)	15 110 (62.1)	64 395 (264.6)
21-30	1 654 423	3772 (22.8)	13 781 (83.3)	14 806 (182.9)	20 458 (242.2)	7570 (45.8)	9614 (58.1)	44 090 (266.5)

^aFor prostate cancer, denominator only included male according to distance from oil refinery: 0-10: 1 089 118; 11-20: 1 188 656; 21-30: 809 518.

^bFor breast cancer, denominator only included female according to distance from oil refinery: 0-10: 1 124 735; 11-20: 1 245 321; 21-30: 844 683.

among persons living within 11-20 miles than either 0-10 or 21-30 miles of an oil refinery.

A total of 829 774 cancer patients were living in Texas from 2001 to 2014. Of those, 283 604 (34.2%) were living within a 30-mile radius of an oil refinery. These cancer patients belonged to 441 zip codes in 99 counties. Characteristics of cancer patients based on proximity to an oil refinery are summarized in [Supplementary Table 3](#) (available online). The age-standardized rates according to cancer type by proximity to an oil refinery are depicted in [Table 2](#), with corresponding heat maps illustrated in [Figure 1](#). We observed an increased rate across all cancer types according to proximity to an oil refinery. We did not observe this trend among the noncancer control fracture cohort.

Adjusted association of proximity to an oil refinery and cancer risk is shown in [Table 3](#). We observed an increased risk of incident cancer diagnosis with proximity to an oil refinery across all cancer types. For example, persons residing within 0-10 (risk ratio [RR] = 1.13, 95% CI = 1.07 to 1.19) and 11-20 (RR = 1.05, 95% CI = 1.00 to 1.11) miles were statistically significantly more likely to be diagnosed with lymphoma than individuals who lived within 21-30 miles of an oil refinery. Older and male vs female individuals were at increased risk of being diagnosed across all nonsex-dependent cancer types (ie, prostate is male only, and breast is predominantly female only). We did not observe any marked racial differences or association with number of oil wells in incident cancer diagnosis across any of the cancers. Cancer rates according to median household income were heterogeneous. However, individuals in the lowest median household income quartile were statistically significantly more likely to be diagnosed with lung cancer when compared with the highest quartile (RR = 1.23, 95% CI = 1.09 to 1.39). We observed a statistically significantly increased risk of cancer associated with smoking across all cancer types with up to twofold increased risk of bladder cancer (RR = 2.15, 95% CI = 1.67 to 2.77) for every 10% increase in smoking-population density. We observed a statistically significantly increased risk of lung, prostate, and breast cancers and lymphoma for every 10% increase in PM_{2.5}. There was no statistically significant difference in incident fracture diagnosis according to any of the distances from an oil refinery.

Differences in stage of cancer at diagnosis according to proximity to an oil refinery was assessed ([Table 4](#)). We found increased risk of localized disease across all cancer types among persons residing 0-10 vs 21-30 miles from an oil refinery. We also found persons residing within 0-10 miles were more likely to be diagnosed with distant metastasis and/or systemic disease than people residing 21-30 miles from an oil refinery. The greatest risk of advanced disease was observed in patients

diagnosed with bladder cancer living within 0-10 vs 21-30 miles (RR = 1.30, 95% CI = 1.02 to 1.65), respectively.

[Supplementary Table 4](#) (available online) provides results from sensitivity analyses. When we limited our analysis to cancer case patients from 2010 to 2014, there was no statistically significant association of proximity to oil refinery and cancer risk except for lung cancer. Our primary study findings persisted in our sensitivity analysis among patients 60 years or older, males (except for bladder cancer), and females. These findings did not persist among patients 20-59 years (except for breast cancer).

Discussion

Cancer is the second most common cause of death in adults aged 25 years or older and is expected to soon be the leading cause of death in the United States ([26,32](#)). Mass industrial production and exposure to potential carcinogenic compounds including pollutants from nearby oil refineries present a public health concern ([33](#)). In our primary analysis, we observed that proximity to an oil refinery was associated with a statistically significantly increased risk of multiple cancer types (bladder, breast, colon, lung, lymphoma, and prostate). Moreover, we found that persons who resided within 0-10 miles had an increased risk of more advanced disease compared with persons residing 21-30 miles from an oil refinery.

Our study revealed several important findings. First, we found an increased risk of all cancer types studied according to proximity to an oil refinery. Prior studies have shown that toxins associated with oil refinery processes may increase the risk of cancer in nearby residents ([7,10, 12-14, 34,35](#)). Hazardous air pollutants known to be emitted from refineries include the highly volatile benzene, toluene, ethylbenzene, and xylene compounds ([2](#)). Many other hydrocarbon pollutants and volatile compounds have been identified near oil refineries, raising the specter of potential serious health risks ([36](#)).

Second, we observed a dose-response relationship between distance and the rate of all cancer sites, because patients living 0-10 miles had the greatest risk of cancer diagnosis followed by 11-20 miles. These findings have important implications because we controlled for known risk factors for cancer including smoking and air pollution. Of note, our findings also controlled for oil well densities, which may contribute to the risk of cancer and alter observed rates ([37](#)). Moreover, the main study findings were consistent in residents 60 years or older, males and females. However, in further sensitivity analyses, we observed non-statistically significant association of proximity to refinery and cancer case patients among cancer case patients from 2010

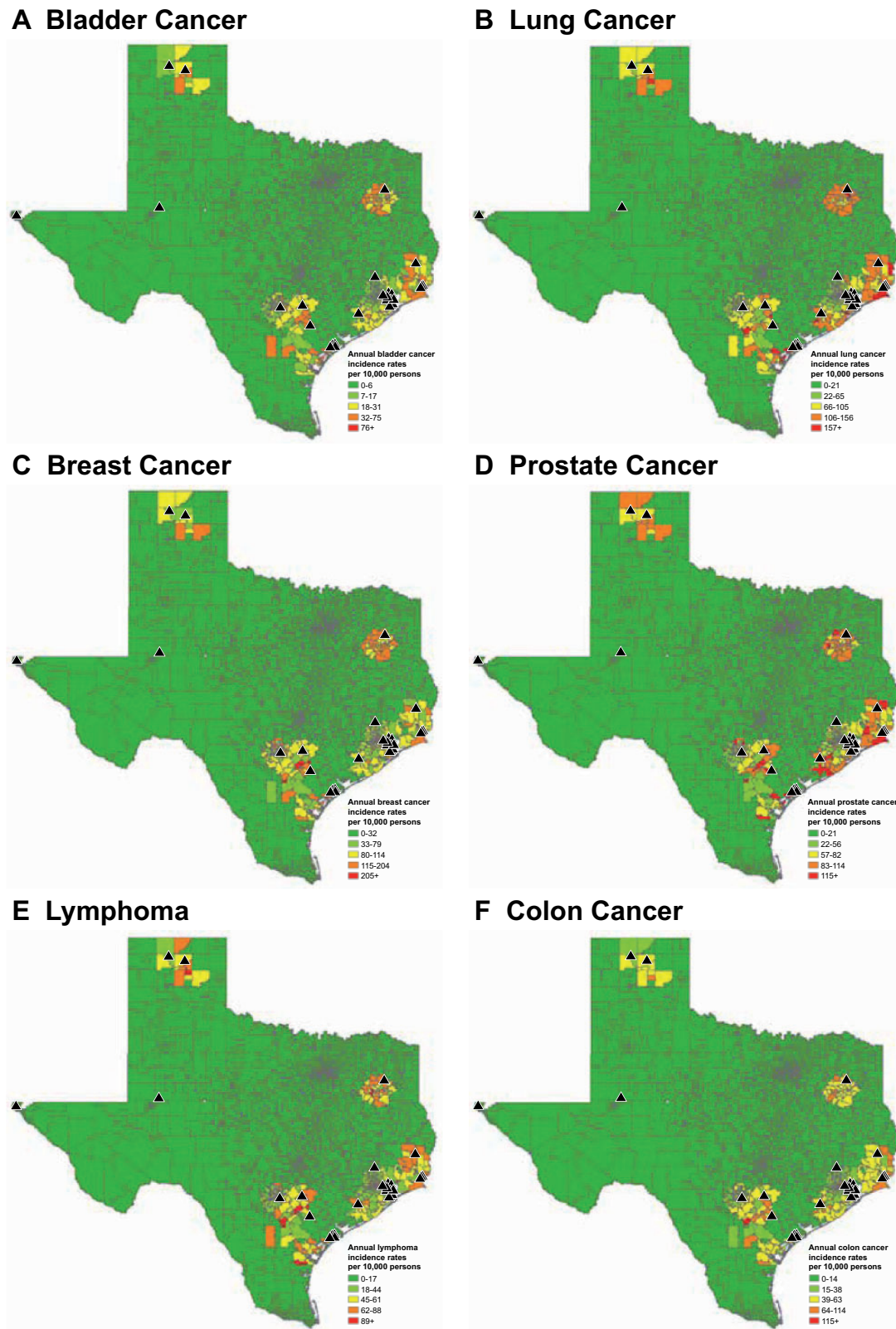


Figure 1. Heat map of cancer rate according to cancer type in Texas in relation to refinery location from 2001 to 2014.

to 2014, which could be because of small number of cancer case patients during that time period. Although we included more 6 million persons living within 30 miles of an oil refinery, our

ability to make a strong causal inference is limited because of a lack of granularity among these “big lens” data. Our findings are hypothesis generating and warrant further individual-level

Table 3. Adjusted association of proximity to an oil refinery and cancer risk

Variable	Risk ratio (95% CI)							Control patients Fracture
	Bladder	Lung	Prostate	Breast	Colon	Lymphoma	Fracture	
Distance from an oil refinery, miles								
0-10	1.07 (1.00 to 1.14)	1.11 (1.04 to 1.20)	1.09 (1.02 to 1.15)	1.13 (1.08 to 1.18)	1.10 (1.04 to 1.17)	1.13 (1.07 to 1.19)	1.01 (0.95 to 1.07)	
11-20	1.03 (0.96 to 1.10)	1.05 (0.98 to 1.13)	1.05 (0.99 to 1.11)	1.06 (1.01 to 1.11)	1.01 (0.95 to 1.07)	1.05 (1.00 to 1.11)	0.96 (0.90 to 1.02)	
21-30	Referent	Referent	Referent	Referent	Referent	Referent	Referent	
P_{trend}^a	.068	.004	.006	<.001	.002	<.001	N/A	
Sex								
Male	3.16 (3.04 to 3.28)	1.43 (1.41 to 1.45)	—	—	1.21 (1.18 to 1.23)	1.33 (1.31 to 1.36)	0.58 (0.58 to 0.59)	
Female	Referent	Referent	—	—	Referent	Referent	Referent	
Age group, y								
20-39	0.04 (0.03 to 0.04)	0.02 (0.02 to 0.02)	0.01 (0.00 to 0.01)	0.08 (0.07 to 0.08)	0.04 (0.04 to 0.05)	0.09 (0.09 to 0.09)	0.02 (0.02 to 0.02)	
40-59	0.16 (0.15 to 0.17)	0.16 (0.16 to 0.16)	0.18 (0.18 to 0.19)	0.59 (0.58 to 0.59)	0.24 (0.24 to 0.25)	0.27 (0.26 to 0.27)	0.07 (0.07 to 0.07)	
≥60	Referent	Referent	Referent	Referent	Referent	Referent	Referent	
Race and ethnicity ^b								
White increase 1%	1.01 (1.00 to 1.01)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.01 (1.01 to 1.01)	
Household income, quartile								
\$0-\$37 506	1.18 (1.04 to 1.33)	1.23 (1.09 to 1.39)	1.08 (0.98 to 1.19)	1.01 (0.93 to 1.08)	1.33 (1.20 to 1.47)	1.17 (1.07 to 1.27)	1.20 (1.08 to 1.32)	
\$37 507-\$46 911	1.27 (1.15 to 1.39)	1.24 (1.12 to 1.36)	1.07 (0.99 to 1.16)	1.04 (0.98 to 1.11)	1.24 (1.14 to 1.34)	1.13 (1.05 to 1.21)	1.21 (1.12 to 1.31)	
\$46 912-\$60 026	1.16 (1.07 to 1.27)	1.09 (0.99 to 1.19)	1.02 (0.95 to 1.10)	0.97 (0.92 to 1.03)	1.14 (1.06 to 1.23)	1.06 (1.00 to 1.13)	1.12 (1.04 to 1.21)	
>\$60 026	Referent	Referent	Referent	Referent	Referent	Referent	Referent	
High school diploma or higher education, quartile								
0-79.7	0.83 (0.75 to 0.93)	1.19 (1.06 to 1.33)	0.73 (0.67 to 0.80)	0.80 (0.74 to 0.85)	1.06 (0.97 to 1.17)	0.87 (0.80 to 0.94)	0.83 (0.76 to 0.92)	
79.8-87.3	0.96 (0.87 to 1.06)	1.30 (1.18 to 1.44)	0.83 (0.76 to 0.90)	0.85 (0.80 to 0.90)	1.12 (1.03 to 1.22)	0.94 (0.87 to 1.00)	0.92 (0.84 to 1.00)	
87.4-92.5	1.07 (0.98 to 1.17)	1.21 (1.1 to 1.33)	0.89 (0.83 to 0.96)	0.91 (0.86 to 0.96)	1.10 (1.02 to 1.19)	1.00 (0.94 to 1.07)	0.99 (0.91 to 1.07)	
>92.5	Referent	Referent	Referent	Referent	Referent	Referent	Referent	
Smoking increase 10% ^c	2.15 (1.67 to 2.77)	1.93 (1.50 to 2.48)	1.44 (1.17 to 1.76)	0.97 (0.83 to 1.14)	1.88 (1.51 to 2.33)	1.48 (1.23 to 1.79)	1.70 (1.38 to 2.09)	
PM2.5 increase 10 units ^d	0.92 (0.40 to 2.11)	1.96 (0.85 to 4.49)	1.36 (0.69 to 2.71)	2.42 (1.43 to 4.10)	0.95 (0.47 to 1.94)	1.08 (0.58 to 2.02)	1.81 (0.91 to 3.62)	
No. of oil wells increase 100 ^e	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	

^aLinear trend was tested through the linear contrast from the multilevel Poisson regression model. CI = confidence interval; N/A = not applicable; PM2.5 = particulate matter 2.5 micrometers or less in size in micrograms per cubic meter.

^bIncrease in proportion of Whites by 1% in each zip code.

^cIncrease in proportion of residents who smoke by 10% in each zip code.

^dIncrease in PM2.5 level by 10 units in each county.

^eIncrease in number of oil wells by 100 in each county. The upper and lower limit of confidence interval had values after fourth decimal.

Table 4. Adjusted association of proximity to oil refinery and cancer risk by cancer stage^a

Cancer type and distance from oil refinery, miles	Adjusted association, RR (95% CI)		
	Localized stage ^b	Regional stage ^c	Distant stage ^d
Bladder cancer			
0-10	1.04 (0.97 to 1.12)	0.98 (0.82 to 1.18)	1.30 (1.02 to 1.65)
11-20	1.05 (0.97 to 1.13)	0.99 (0.82 to 1.18)	0.98 (0.77 to 1.26)
21-30	Referent	Referent	Referent
Lung cancer			
0-10	1.10 (1.00 to 1.20)	1.12 (1.02 to 1.23)	1.14 (1.05 to 1.23)
11-20	1.05 (0.95 to 1.15)	1.07 (0.97 to 1.17)	1.06 (0.98 to 1.15)
21-30	Referent	Referent	Referent
Prostate cancer			
0-10	1.11 (1.04 to 1.19)	1.06 (0.96 to 1.16)	1.24 (1.09 to 1.41)
11-20	1.11 (1.04 to 1.18)	1.10 (1.01 to 1.20)	1.01 (0.88 to 1.15)
21-30	Referent	Referent	Referent
Breast cancer			
0-10	1.11 (1.06 to 1.17)	1.13 (1.07 to 1.20)	1.19 (1.08 to 1.32)
11-20	1.06 (1.01 to 1.12)	1.05 (0.99 to 1.12)	1.07 (0.97 to 1.19)
21-30	Referent	Referent	Referent
Colon cancer			
0-10	1.12 (1.04 to 1.21)	1.10 (1.02 to 1.19)	1.13 (1.03 to 1.24)
11-20	1.01 (0.94 to 1.10)	1.04 (0.97 to 1.12)	1.01 (0.92 to 1.11)
21-30	Referent	Referent	Referent
Lymphoma			
0-10	1.17 (1.07 to 1.28)	1.08 (0.96 to 1.21)	1.17 (1.10 to 1.24)
11-20	1.18 (1.08 to 1.28)	1.03 (0.92 to 1.16)	1.06 (1.00 to 1.12)
21-30	Referent	Referent	Referent

^aModel controlled for age, sex, race and ethnicity, income, education, smoking, oil wells, and PM2.5. CI = confidence interval; RR = risk ratio.

^bLocalized includes in situ and localized cancer.

^cRegional includes regional by direct extension only, regional lymph nodes only, regional (direct extension and regional lymph nodes) and regional not specified.

^dDistant includes distant metastasis and/or systemic disease.

investigation including epidemiological fieldwork with biological interpretation.

Third, we found increased risk of regional and more advanced metastatic disease according to proximity to an oil refinery. These findings persisted across all cancer types and highlight not only the increased risk of cancer but also the burden of the disease. We controlled for socioeconomic determinants including median household income, and these findings continued to persist, with persons living within 0-10 miles having increased risk of metastatic disease than individuals living more than 20 miles from an oil refinery. Given the increased rate in advanced malignancies in all cancer sites that we observed in relation to refinery proximity, and especially in the setting of low socioeconomic status, we highlight the importance of screening and intervention efforts for an increasingly at-risk population living in proximity to PIC (38).

Fourth, we did not observe an association between disease and proximity to an oil refinery among our negative control outcome, fracture diagnosis. We were unable to determine an association between fracture diagnosis and proximity to an oil refinery. These findings among our control cohort are salient because they refute the potential impact of an unmeasured confounder on the observed association of cancer risk and proximity to an oil refinery.

Our findings must be considered within the context of the study design. First, we used a cross-sectional study design that lacked individual-level data including exposure information, which has inherent selection bias. Further sensitivity analyses

among older patients and according to sex had similar findings regarding risk of cancer diagnosis and proximity to an oil refinery. However, these observations need to be balanced with the non-statistically significant association of proximity to a refinery among cancer case patients from 2010 to 2014. Although we attribute that this finding could be due to small number of cancer case patients during that time period, we cannot exclude potential uncontrolled confounders contributing to that observation. Second, we did not capture other potential confounders including access to care, radiation exposure, and occupational exposure. Although we attempted to control for the confounders including smoking and air pollution (PM2.5), we acknowledge the limitations in using county-level data. Fourth, we did not capture duration and prior residence history, which may influence our results. Fifth, socioeconomic factors are important confounders and should be controlled. Because of the nature of the data and analysis, we could only control for zip code-level income as a proxy for socioeconomic status. Finally, we did not assess the size and capacity of oil refineries included or incidents of accidental discharges and prior explosions.

In conclusion, using a population-based cohort in a state hosting the largest number of oil refineries, proximity to an oil refinery was associated with an increased risk of multiple cancer types. We also observed statistically significantly increased risk of metastatic disease associated with proximity to an oil refinery. These findings support the need for further individual-level investigation into risk of carcinogenesis linked to proximity to an oil refinery.

Funding

This study was conducted with the support of a Department of Defense Peer Reviewed Cancer Research Program Career Development Award (W81XWH1710576) (SBW).

Notes

Role of the funder: The funder had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

Disclosures: The authors declare no potential conflicts of interest.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the DSHS or HHSC.

Acknowledgments: This study used the Texas Cancer Registry databases. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the Department of Social Health Services and Texas Health and Human Services Commission and the TCR program tumor registry in the creation of the TCR database.

Author contributions: Conceptualization- Stephen B. Williams; Yong Shan; Preston S. Kerr; Hemalkumar B. Mehta. Data curation- Stephen B. Williams; Yong Shan; Preston S. Kerr; Ikenna Okereke; Hemalkumar B. Mehta. Formal analysis- Yong Shan; Yong-Fang Kuo; Hemalkumar B. Mehta; Funding acquisition- Stephen B. Williams. Investigation- Stephen B. Williams; Yong Shan; Usama Jazzar; Preston S. Kerr; Ikenna Okereke; Jacques G. Baillargeon; Yong-Fang Kuo; Hemalkumar B. Mehta. Methodology- Stephen B. Williams; Yong Shan; Usama Jazzar; Preston S. Kerr; Ikenna Okereke; Jacques G. Baillargeon; Yong-Fang Kuo; Hemalkumar B. Mehta. Project administration- Stephen B. Williams; Douglas S. Tyler; Yong-Fang Kuo; Hemalkumar B. Mehta. Resources- Stephen B. Williams; Yong-Fang Kuo; Hemalkumar B. Mehta. Software- Yong Shan, Douglas S. Tyler; Stephen B. Williams; Supervision- Douglas S. Tyler, Stephen B. Williams; Validation- Stephen B. Williams; Yong Shan; Jacques G. Baillargeon; Yong-Fang Kuo; Hemalkumar B. Mehta. Visualization- Stephen B. Williams; Yong Shan; Usama Jazzar; Preston S. Kerr; Ikenna Okereke; Jacques G. Baillargeon; Yong-Fang Kuo; Hemalkumar B. Mehta. Writing - original draft- Stephen B. Williams; Yong Shan; Usama Jazzar; Preston S. Kerr; Ikenna Okereke; Jacques G. Baillargeon; Yong-Fang Kuo; Hemalkumar B. Mehta. Writing - review & editing- Stephen B. Williams; Yong Shan; Usama Jazzar; Preston S. Kerr; Ikenna Okereke; V. Suzanne Klimberg; Douglas S. Tyler; Nagireddy Putluri; David S. Lopez; John D. Prochaska; Cornelis Elferink; Jacques G. Baillargeon; Yong-Fang Kuo; Hemalkumar B. Mehta

Data Availability

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Texas Department of State Health Services (DSHS) or Health and Human Services Center (HHSC). This study used the Texas Cancer Registry (TCR) databases. The interpretation and

reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the DSHS and HHSC; and the TCR program tumor registry in the creation of the TCR database. The data underlying this article were provided by the TCR under license / by permission. Data will be shared on request to the corresponding author with permission of the TCR.

References

1. US Energy Information Administration. Petroleum & other liquids: refinery capacity report. <https://www.eia.gov/petroleum/refinerycapacity/>. Accessed October 1, 2018.
2. Fleischman L, Banks J, Graham J. Fossil fumes: a public health analysis of toxic air pollution from the oil and gas industry. <https://www.catf.us/resource/fossil-fumes-public-health-analysis/>. Accessed October 1, 2018.
3. Foster GW, Lipinski DW, Schneider BS, et al. H.R. 1165: expanding transparency of information and safeguarding toxics act of 2019. <https://www.congress.gov/116/bills/hr1165/BILLS-116hr1165ih.pdf>. Accessed October 1, 2018.
4. US Environmental Protection Agency (EPA). Oil and natural gas sector: emission standards for new, reconstructed, and modified sources reconsideration. <https://www.federalregister.gov/documents/2018/10/15/2018-20961/oil-and-natural-gas-sector-emission-standards-for-new-reconstructed-and-modified-sources>. Accessed October 1, 2018.
5. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Benzene. Lyon, France: International Agency for Research on Cancer; 2018 (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 120). <https://www.ncbi.nlm.nih.gov/books/NBK550157/>.
6. Moorthy B, Chu C, Carlin DJ. Polycyclic aromatic hydrocarbons: from metabolism to lung cancer. *Toxicol Sci*. 2015;145(1):5-15.
7. Simonsen N, Scribner R, Su LJ, et al. Environmental exposure to emissions from petrochemical sites and lung cancer: the lower Mississippi Interagency Cancer Study. *J Environ Public Health*. 2010;2010:1-9.
8. Axelsson G, Barregard L, Holmberg E, et al. Cancer incidence in a petrochemical industry area in Sweden. *Sci Total Environ*. 2010;408(20):4482-4487.
9. Wilkinson P, Thakrar B, Walls P, et al. Lymphohaematopoietic malignancy around all industrial complexes that include major oil refineries in Great Britain. *Occup Environ Med*. 1999;56(9):577-580.
10. Weng HH, Tsai SS, Chiu HF, et al. Association of childhood leukemia with residential exposure to petrochemical air pollution in Taiwan. *Inhal Toxicol*. 2008;20(1):31-36.
11. Yang CY, Cheng MF, Chiu JF, et al. Female lung cancer and petrochemical air pollution in Taiwan. *Arch Environ Health*. 1999;54(3):180-185.
12. Yu CL, Wang SF, Pan PC, et al. Residential exposure to petrochemicals and the risk of leukemia: using geographic information system tools to estimate individual-level residential exposure. *Am J Epidemiol*. 2006;164(3):200-207.
13. Lin CK, Hsu YT, Christiani DC, et al. Risks and burden of lung cancer incidence for residential petrochemical industrial complexes: a meta-analysis and application. *Environ Int*. 2018;121:404-414.
14. Sans S, Elliott P, Kleinschmidt I, et al. Cancer incidence and mortality near the Baglan Bay petrochemical works, South Wales. *Occup Environ Med*. 1995; 52(4):217-224.
15. Texas Department of State Health Services: Epidemiology and Surveillance Branch. Texas Cancer Registry (TCR). <https://www.dshs.state.tx.us/tcr/home.aspx>. Accessed October 1, 2018.
16. Liu Z, Zhang Y, Franzin L, et al. Trends and variations in breast and colorectal cancer incidence from 1995 to 2011: a comparative study between Texas Cancer Registry and National Cancer Institute's Surveillance, Epidemiology and End Results data. *Int J Oncol*. 2015;46(4):1819-1826.
17. United States Census Bureau. Population and housing unit estimates. <https://www.census.gov/programs-surveys/popest.html>. Accessed October 1, 2018.
18. County Health Rankings and Roadmaps. Texas rankings data. <https://www.countyhealthrankings.org/rankings/data/TX>. Accessed October 1, 2018.
19. Hendryx M, Ahern MM, Zullig KJ. Improving the environmental quality component of the county health rankings model. *Am J Public Health*. 2013;103(4): 727-732.
20. Remington PL, Catlin BB, Gennuso KP. The county health rankings: rationale and methods. *Popul Health Metr*. 2015;13(11):1-12.
21. Oliver TR. Population health rankings as policy indicators and performance measures. *Prev Chronic Dis*. 2010;7(5):A101.
22. Jerrett M, Arain A, Kanaroglou P, et al. A review and evaluation of intraurban air pollution exposure models. *J Expo Sci Environ Epidemiol*. 2005;15(2):185-204.
23. Briggs D. The role of GIS: coping with space (and time) in air pollution exposure assessment. *J Toxicol Environ Health A*. 2005;68(13-14):1243-1261.
24. Singh Z, Chadha P. Textile industry and occupational cancer. *J Occup Med Toxicol*. 2016;11(1):39-39.
25. Ayuso-Álvarez A, García-Pérez J, Triviño-Juárez J-M, et al. Association between proximity to industrial chemical installations and cancer mortality in Spain. *Environ Pollut*. 2020;260:113869.
26. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA A Cancer J Clin*. 2019; 69(1):7-34.

27. Blot WJ, Brinton LA, Fraumeni JF, et al. Cancer mortality in U.S. counties with petroleum industries. *Science*. 1977;198(4312):51–53.
28. Texas Commission on Environmental Quality. PM 2.5 data: soot, dust, smoke (particulate matter). <https://www.tceq.texas.gov/agency/data/lookup-data/pm25.html>. Accessed October 1, 2018.
29. Railroad Commission of Texas. Well distribution by county - well counts. <https://www.rrc.state.tx.us/oil-gas/research-and-statistics/well-information/well-distribution-by-county-well-counts/>. Accessed October 1, 2018.
30. Cheung YB. Zero-inflated models for regression analysis of count data: a study of growth and development. *Statist Med*. 2002;21(10):1461–1469.
31. Ridout M, Hinde J, Demétrio CG. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*. 2001;57(1):219–223.
32. National Center for Health Statistics. Health, United States, 2017: with special feature on mortality. <https://www.ncbi.nlm.nih.gov/pubmed/30702833>.
33. Clapp RW, Jacobs MM, Loechler EL. Environmental and occupational causes of cancer: new evidence 2005-2007. *Rev Environ Health*. 2008;23(1):1–37.
34. Linos A, Blair A, Gibson RW, et al. Leukemia and non-Hodgkin's lymphoma and residential proximity to industrial plants. *Arch Environ Health*. 1991;46(2):70–74.
35. Johnson KC, Pan S, Fry R, et al. Residential proximity to industrial plants and non-Hodgkin lymphoma. *Epidemiology*. 2003;14(6):687–693.
36. Sansom GT, Kirsch KR, Stone KW, et al. Domestic exposures to polycyclic aromatic hydrocarbons in a Houston, Texas, Environmental Justice Neighborhood. *Environ Justice*. 2018;11(5):183–191.
37. McKenzie LM, Allshouse WB, Byers TE, et al. Childhood hematologic cancer and residential proximity to oil and gas development. *PLoS One*. 2017;12(2):e0170423.
38. Meyskens FL, Mukhtar H, Rock CL, et al. Cancer prevention: obstacles, challenges and the road ahead. *J Natl Cancer Inst*. 2016;108(2):1–8.