

Supplementary Table 1. Data sources used in this study

Datasets	Year	Weblinks	Why we choose this data?
Texas Cancer Registry	2001-2014	n/a	To get cancer cases
2010-2014 American Community Survey 5-year estimates	2010-2014	https://www.census.gov/data/developers/data-sets/acs-5year.2014.html	Age, sex, race, income and education at ZIP level (Income and education information became available from 2010 Census. It was not available in 2000 Census data).
County health rankings & roadmaps	2016	https://www.countyhealthrankings.org/app/texas/2016/measure/factors/9/map	Smoking at county level estimated from 2014 data (More than 67% missing data on smoking for previous years)
County health rankings & roadmaps PM2.5	2013	https://www.countyhealthrankings.org/app/texas/2019/measure/factors/125/map	2013 data captures average daily PM2.5 from 2008 data
Railroad commission of Texas	2003	https://www.rrc.state.tx.us/media/2122/oilwellct0903.pdf	Number of oil wells from 2003

Supplementary Tale 2. List of 28 oil refineries

Refinery Location	Corporate Owner	Zip Code	First operational	Years Inactive
Baytown	Exxon Mobil Corporation	77520	1920	Na
Beaumont	Exxon Mobil Corporation	77701	1903	Na
Big Spring	Alon USA Energy Inc.	79720	2000	Na
Borger	WRB Refining, LLC	79007	1926	Na
Channelview	Equistar Chemicals	77049	1957	Na
Corpus Christi	BTB Refining	78409	2007	2013-Present
Corpus Christi	Citgo Petroleum Corporation	78401	1935	Na
Corpus Christi	Flint Hills Resources	78409	1981	Na
Corpus Christi	Valero Energy Corporation	78407	1983	Na
Deer Park	Deer Park Refining	77536	1929	Na
El Paso	Western Refining Inc.	79905	1997	Na
Galena Park	Kinder Morgan	77547	2015	Na
Galveston Bay of Texas City	Marathon Petroleum	77590	1934	Na
Houston	Houston Refining LP	77017	1918	Na
Houston	Valero Energy Corporation	77012	1942	Na
Sunray	Valero Energy Corporation	79086	1933	Na
Nixon	Lazarus Energy, LLC	78140	1980	1993-2011
Pasadena	Pasadena Refining	77506	1919	Na
Port Arthur	Motiva Enterprises, LLC	77640	1902	Na
Port Arthur	Total Petrochemicals and Refining	77619	1936	Na
Port Arthur	Valero Energy Corporation	77640	1901	Na
San Antonio	Calumet Specialty Products Partners	78223	1955	Na
Silsbee	South Hampton Resources	77656	1955	Na
Sweeny	Phillips 66	77463	1942	Na
Texas City	Marathon Petroleum	77590	1931	Na
Texas City	Valero Energy Corporation	77590	1908	Na
Three Rivers	Valero Energy	78071	1974	Na

	Corporation			
Tyler	Delek US Holdings, Inc.	75702	1923	Na

Supplementary Table 3. Characteristics of cancer patients based on distance from an oil refinery

Characteristics	Distance from oil refineries			P-value
	0-10 miles	11-20 miles	21-30 miles	
Bladder cancer, N =14924				
Age, mean (SD)	71.0 (12.3)	70.5 (12.2)	69.6 (12.2)	<0.001
Gender, N (%)				<0.001
Male	4016 (73.1)	4255 (76.3)	2868 (76.9)	
Female	1476 (26.9)	1327 (23.7)	864 (23.1)	
Race/ethnicity, N (%)				<0.001
White	4671 (85.1)	5018 (89.9)	3410 (91.4)	
Black	729 (13.3)	370 (6.6)	165 (4.4)	
Other	92 (1.7)	194 (3.5)	157 (4.2)	
Cancer stage, N (%)				<0.001
Localized	4108 (74.8)	4390 (78.6)	2916 (78.1)	
Regional	400 (7.3)	400 (7.2)	266 (7.1)	
Distant	266 (4.8)	204 (3.7)	135 (3.6)	
Unstaged	718 (13.1)	588 (10.5)	415 (11.1)	
Lung cancer, N =58067				
Age, mean (SD)	68.8(11.5)	68.9(11.3)	68.3(11.4)	<0.001
Gender, N (%)				<0.001
Male	13116(56.4)	11201(53.8)	7477(54.4)	
Female	10114(43.5)	9606(46.1)	6258(45.5)	
Race/ethnicity, N (%)				<0.001
White	17525(75.4)	17458(83.9)	12141(88.3)	
Black	5337(22.9)	2646(12.7)	1122(8.1)	
Other	368(1.5)	703(3.3)	472(3.4)	
Cancer stage, N (%)				0.026
Localized	3768(16.2)	3566(17.1)	2404(17.5)	
Regional	4851(20.8)	4331(20.8)	2886(21)	
Distant	10645(45.8)	9412(45.2)	6102(44.4)	
Unstaged	3966(17.0)	3498(16.8)	2343(17.0)	
Prostate cancer, N =57603				
Age, mean (SD)	67.5(10.1)	66.4(9.9)	65.7(9.9)	<0.001
Gender, N (%)				<0.001
Male	20515(100)	21614(100)	14572(100)	
Female				
Race/ethnicity, N (%)				<0.001
White	14987(73)	17221(79.6)	12258(84.1)	
Black	5002(24.3)	3433(15.8)	1618(11.1)	
Other	526(2.5)	960(4.4)	696(4.7)	

Characteristics	Distance from oil refineries			P-value
	0-10 miles	11-20 miles	21-30 miles	
Cancer stage, N (%)				<0.001
Localized	14002(68.2)	15399(71.2)	10159(69.7)	
Regional	1789(8.7)	2105(9.7)	1444(9.9)	
Distant	1080(5.2)	856(3.9)	547(3.7)	
Unstaged	3644(17.7)	3254(15)	2422(16.6)	
Breast cancer, N =80685				
Age, mean (SD)	61(13.8)	59.7(13.5)	58.8(13.2)	<0.001
Gender, N (%)				
Male				
Female	28733(100)	31113(100)	20358(100)	
Race/ethnicity, N (%)				<0.001
White	22161(77.1)	25361(81.5)	17077(83.8)	
Black	5895(20.5)	4371(14)	2004(9.8)	
Other	677(2.3)	1381(4.4)	1277(6.2)	
Cancer stage, N (%)				<0.001
Localized	17720(61.6)	20303(65.2)	13214(64.9)	
Regional	7189(25)	7122(22.8)	4662(22.9)	
Distant	1636(5.6)	1497(4.8)	938(4.6)	
Unstaged	2188(7.6)	2191(7)	1544(7.5)	
Colon cancer, N =32283				
Age, mean (SD)	67.8(13.7)	66.5(13.9)	65.9(14)	<0.001
Gender, N (%)				0.094
Male	6516(50.4)	5939(51.1)	3921(51.9)	
Female	6411(49.5)	5681(48.8)	3623(48)	
Race/ethnicity, N (%)				<0.001
White	9476(73.3)	9390(80.8)	6310(83.6)	
Black	3215(24.8)	1784(15.3)	796(10.5)	
Other	236(1.8)	446(3.8)	438(5.8)	
Cancer stage, N (%)				0.064
Localized	4623(35.7)	4118(35.4)	2686(35.6)	
Regional	4241(32.8)	3974(34.1)	2495(33)	
Distant	2603(20.1)	2227(19.1)	1446(19.1)	
Unstaged	1460(11.2)	1301(11.1)	917(12.1)	
Lymphoma, N =40029				
Age, mean (SD)	63.8(16.6)	63(16.7)	62.7(16.1)	<0.001
Gender, N (%)				0.005
Male	7988(53.1)	8068(54.2)	5197(55.2)	
Female	7035(46.8)	6793(45.7)	4203(44.7)	
Race/ethnicity, N (%)				<0.001
White	11988(79.7)	12479(83.9)	8126(86.4)	

Characteristics	Distance from oil refineries			P-value
	0-10 miles	11-20 miles	21-30 miles	
Black	2702(17.9)	1808(12.1)	821(8.7)	
Other	333(2.2)	574(3.8)	453(4.8)	
Cancer stage, N (%)				<0.001
Localized	1705(11.3)	1886(12.6)	1090(11.5)	
Regional	1005(6.6)	1040(6.9)	654(6.9)	
Distant	10302(68.5)	9838(66.2)	6180(65.7)	
Unstaged	2011(13.3)	2097(14.1)	1476(15.7)	

* Localized includes in situ and localized cancer.

** Regional includes regional by direct extension only, regional lymph nodes only, Regional (direct extension and regional lymph nodes) and regional not specified

*** Distant includes distant metastasis and/or systemic disease

Supplementary Table 4. Adjusted association of proximity to an oil refinery and cancer risk in elderly (65 years and older), males, and female population, and restricting to cancer cases in 2010-2014

Distance from Oil Refinery (Miles)	Risk Ratio for Each Cancer Site (95% CI)					
	Bladder	Lung	Prostate	Breast	Colon	Lymphoma
	Cancer cases and ACS population from 2010-2014					
0-10	1.02 (0.94-1.11)	1.15 (1.00-1.33)	1.02 (0.98-1.06)	1.02 (0.96-1.09)	1.03 (0.96-1.11)	1.05 (0.99-1.11)
11-20	0.93 (0.86-1.01)	1.02 (0.88-1.18)	0.99 (0.95-1.04)	1.00 (0.94-1.07)	0.97 (0.90-1.04)	0.98 (0.92-1.03)
21-30	Ref	Ref	Ref	Ref	Ref	Ref
	Older adults (60 years or older)					
0-10	1.10 (1.05-1.57)	1.07 (1.05-1.10)	1.09 (1.06-1.11)	1.12 (1.09-1.15)	1.14 (1.10-1.18)	1.13 (1.09-1.17)
11-20	1.05 (1.00-1.10)	1.01 (0.99-1.04)	1.07 (1.04-1.10)	1.09 (1.07-1.12)	1.02 (0.98-1.06)	1.06 (1.03-1.10)
21-30	Ref	Ref	Ref	Ref	Ref	Ref
	Younger adults (20 to 59 years).					
0-10	1.08 (0.93-1.26)	1.09 (0.96-1.25)	1.03 (0.93-1.15)	1.10 (1.02-1.19)	0.93 (0.83-1.05)	1.07 (0.98-1.17)
11-20	1.05 (0.91-1.22)	0.99 (0.87-1.13)	1.03 (0.92-1.14)	1.03 (0.96-1.11)	0.90 (0.80-1.00)	1.03 (0.94-1.12)
21-30	Ref	Ref	Ref	Ref	Ref	Ref
	Male					
0-10	1.05 (0.97-1.13)	1.13 (1.05-1.22)	--	--	1.08 (1.01-1.16)	1.14 (1.07-1.20)
11-20	1.03 (0.96-1.11)	1.04 (0.96-1.12)	--	--	0.99 (0.92-1.06)	1.06 (1.00-1.12)
21-30	Ref	Ref	Ref	Ref	Ref	Ref
	Female					
0-10	1.13 (1.02-1.26)	1.09 (1.01-1.18)	--	--	1.12 (1.04-1.20)	1.14 (1.07-1.21)
11-20	1.03 (0.92-1.15)	1.07 (0.99-1.66)	--	--	1.02 (0.95-1.10)	1.06 (1.00-1.12)
21-30	Ref	Ref	Ref	Ref	Ref	Ref

Model included age, sex, race/ethnicity, income, smoking, oil wells, and PM2.5