



Assessing residential socioeconomic factors associated with pollutant releases using EPA's Toxic Release Inventory

Amanda T. Charette^{1,2} · Mary B. Collins^{1,3} · Jaime E. Mirowsky^{1,2} 

Accepted: 6 January 2021 / Published online: 25 January 2021
© AESS 2021

Abstract

There is a large body of literature showing that minorities and people living in low-income households live disproportionately close to polluting industrial facilities across the United States. However, only limited work of this nature has been conducted in Upstate New York. In this study, we utilized hierarchical clustering to create seven residential clusters from four Upstate New York counties; each cluster was then spatially linked to the locations of the polluting facilities and the quantity of pollutants released. The largest numbers of facilities and the highest quantities of releases were located in two clusters described as primarily working class. The lowest numbers of facilities were located in the two clusters representing neighborhoods that were the most economically deprived and the most wealthy and educated. These findings suggest that, in addition to race and class as predictors of community-level contamination, other metrics of socioeconomic status might help clarify the complex landscape of environmental inequity.

Keywords Toxic Release Inventory (TRI) · Hierarchical clustering · Socioeconomic · Environmental justice

Introduction

On the night of December 2, 1984, a catastrophic industrial disaster occurred in Bhopal, India (United States Environmental Protection Agency 2019a). A gas leak of methylisocyanate occurred at the Union Carbide India Limited pesticide plant causing serious injury or death to over 2,000 people (United States Environmental Protection Agency 2019a). Two years later, in response to that event, the United States (US) passed the Emergency Planning and Community Right-to-Know Act (EPCRA). EPCRA requires federal, state, and local governments; tribes; and industries to have emergency plans in place in case an accident, such as a chemical release

or leak, occurs. Within EPCRA, Section 313 established the Toxic Release Inventory (TRI) program, which tracks the management of over 650 toxic chemicals across the US (United States Environmental Protection Agency 2019b); the facilities that release these chemicals are known as TRI facilities. Information about TRI facilities, including where they are located and the quantity and types of chemicals they release, are publicly available using the TRI Explorer database managed by the US Environmental Protection Agency (United States Environmental Protection Agency 2020a).

It was not until later in the 1980s that a major connection was made between the location of polluting facilities and the neighborhoods they were located in, when one of the first major pieces of literature on this topic was published (Commission for Racial Justice 1987). In this work, researchers found that facilities that were treating, storing, and disposing of hazardous waste were disproportionately located near minority communities. The unequal patterns displayed in this study helped galvanize the environmental justice movement in the US. This movement, which is still ongoing, strives for the fair treatment and meaningful involvement of all people, regardless of race, color, national origin, or income, in the development and implementation of any and all environmental laws and regulations (United States Environmental Protection Agency 2019c).

✉ Jaime E. Mirowsky
jmirowsk@esf.edu

¹ Center for Environmental Medicine and Informatics, State University of New York College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, USA

² Department of Chemistry, State University of New York College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, USA

³ Department of Environmental Studies, State University of New York College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, USA

Now, more than 30 years after that initial report was published, clear signals of environmental injustice are still being observed (Bullard et al. 2007). However, since that first publication, other studies have shown that additional socioeconomic status (SES) variables also may be influential in the siting of polluting facilities, including, but not limited to, occupation, housing, and education level (Boer et al. 1997; Williams 2008; Johnson et al. 2016). To complicate this matter, the choice of SES variables has not been well defined by researchers, and there is no consensus as to the best SES variables to be used for environmental justice research (Messer et al. 2006; Mirowsky et al. 2017). Differences in the geographic scale (i.e., census tracts vs. block groups) used also fluctuates between studies, and it is likely that the geographic scale chosen influences not only the SES factors being examined, but also the final outcomes observed (Perlin et al. 1995; Cutter et al. 1996; Sicotte 2010). Further, as the results appear to differ based on the geographic region being studied, it is also highly plausible that where the study is being conducted also impacts the decisions being made as to where these polluting facilities are located. In a literature review performed by Sicotte (2010), the author stated that similar or identical methodologies in environmental justice studies may yield different patterns of inequality when applied to different metropolitan areas across the United States. For example, one study in Phoenix, Arizona, US, found that the number of hazardous waste sites increased with the population percentages of African Americans and Latinos and decreased with household income and the percent population of Whites (Bolin et al. 2002). Another study in New York found that

percent minority was positively associated with the presence of hazards in Brooklyn, Queens, and the Bronx, but saw a negative association in Manhattan (Fricker Jr. and Hengartner 2001).

To date, very little environmental justice work, as it relates to the location of polluting facilities, has been conducted in the more suburban and rural areas of NY state. For this reason, a deeper exploration into this understudied area is needed. Thus, the main objective of our study was to determine whether certain community-level characteristics were associated with the largest and smallest number of facilities and chemical releases in Upstate New York. The quantity of chemical releases is an important and complimentary factor to consider for this type of work, as some facilities have been shown to release larger amounts of polluting chemicals than others (Ash and Fetter 2002; Collins et al. 2020). This interdisciplinary work will add to the literature on how certain communities may be more vulnerable in experiencing adverse health effects from polluting facilities and what characteristics make up those vulnerable communities.

Materials and methods

Study location

The counties chosen for this study were Erie, Monroe, Onondaga, and Albany Counties, New York. Each county contains a major Upstate New York city, and they are all located within 300 miles of one another (Fig. 1). The population characteristics for the four counties can be seen in Table 1.

Fig. 1 Map of New York State, with (from west to east) Erie, Monroe, Onondaga, and Albany Counties highlighted in yellow

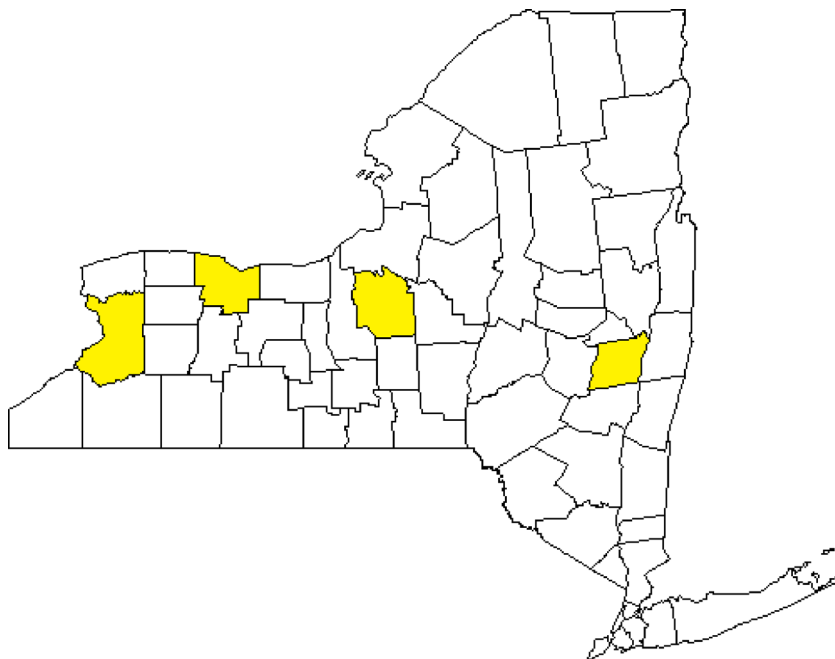


Table 1 Characteristics of four major Upstate New York counties from 2000 (United States Census Bureau 2019)

County name	Total population	Major city	Major city population (number of persons)
Erie	941,498	Buffalo	292,648
Monroe	731,896	Rochester	219,766
Onondaga	458,336	Syracuse	147,326
Albany	294,540	Albany	95,658

Defining residential clusters using US Census data

The data used to create the unique residential clusters were obtained from the 2000 decennial US Census at the block group level using American FactFinder. The year 2000 was chosen because information on the specific variables being studied for this project was not available for 2010. The nine Census variables chosen for this work were categorized into seven sub-categories: education, wealth, income, race/ethnicity, employment, housing, and land use (Mirowsky et al. 2017) (Table 2).

In this study, neighborhood-level SES was assessed using Census variables that were deemed most important given our previous knowledge of the study areas as well as the prior literature. For example, a neighborhood’s education level has been a factor used in multiple studies assessing area-level SES and health (Curry et al. 1993; Evans and Marcynyszyn 2004; Mirowsky et al. 2017; Roberts 1997). The area chosen for this study includes many colleges and universities, but also high baseline levels of poverty (World Population Review 2020); thus, having a high school degree or greater was the variable we selected to represent the Education sub-category and living below the poverty line was the variable we selected to represent the Income sub-category. Further, the volume of universities and colleges can tie into the type of housing that would be expected in the area, as many university or college areas contain rental properties (Mirowsky et al. 2017). Therefore, for our housing category, looking at owner-occupied housing was selected (Arora and

Cason 1999; Mirowsky et al. 2017; Pastor Jr. et al. 2001; Wolverton 2009) as opposed to median home value. Through the literature, it has been speculated that if firms are paying out compensation to all individual members of a community for placing a facility within said community, they will look at the population density of the neighborhood, as well as the amount of houses that are unoccupied (Wolverton 2009). Therefore, we also looked at levels of vacant housing in our study area to account for this possible phenomenon (Mirowsky et al. 2017; Wolverton 2009). Other metrics such as single-parent households (Evans and Marcynyszyn 2004; Mirowsky et al. 2017; Roberts 1997), race (Curry et al. 1993; Evans and Marcynyszyn 2004; Mirowsky et al. 2017), percent unemployment (Curry et al. 1993; Mirowsky et al. 2017), and percent urban (Curry et al. 1993; Evans and Marcynyszyn 2004; Mirowsky et al. 2017) have been used extensively in the literature by other researchers when assessing R-SES, and these were also selected for that reason.

The race/ethnicity sub-category includes the population identifying as any race/ethnicity other than non-Hispanic White. Non-managerial positions were defined as the percentage of employed civilians, over the age of 16 years, regardless of sex, working in the following occupations: service; sales and office; farming, fishing, and forestry; construction, extraction, and maintenance; and production, transportation, and material moving. The percentage of single-parent housing was the number of male or female only (no spouse present) family households in owner- and renter-occupied housing divided by the total number of

Table 2 US Census variables used to form the clusters

Category	Variable details
Education	Percent of the population that has obtained at least a high school diploma
Wealth	Percent of the population in owner-occupied housing
Income	Percent of the population with income below the poverty line
Race/ethnicity	Percent of the population not identifying as White or non-Hispanic White
Employment	Percent of the population unemployed
Housing	Percent of the population in non-managerial positions
	Percent of the population living in single-parent housing
Land-Use	Percent of vacant housing
	Percent urban environment

owner- and renter-occupied housing units. All variables were expressed as percentages.

Hierarchical clustering

Hierarchical clustering analysis was done using Ward's hierarchical clustering method (Ward Jr. 1963) using the nine Census variables from each block group spread across all four counties. Ward's method uses a bottom-up approach to look for similarities in a group of observations, and it was chosen for this analysis because the pooled within group sum of squares is minimized (Mirowsky et al. 2017). To assist in determining the optimal number of clusters, the Friedman method was used (Friedman and Rubin 1967).

By creating these residential clusters, we utilized what Mohai and Saha (2007) refer to as the "unit-hazard coincidence" method. This approach involves selecting a geographic unit, such as a block group, and identifying which of these units contain a chosen hazard, such as a TRI facility. We choose this approach because it is one of the most widely used approaches for addressing demographic disparities in the distribution of environmental hazards, such as TRI facilities, and has been used in both local- (Hill et al. 2018) and national-level studies (Commission for Racial Justice 1987; Bolin et al. 2002; Wilson et al. 2012).

Toxic Release Inventory facilities

The location of every TRI facility in our study area from the year 2000 was obtained from TRI Explorer and geocoded into each cluster (United States Environmental Protection Agency 2020a). TRI Explorer is a tool managed by the EPA to share information regarding the location, releases, and treatment options used by TRI reporting facilities. The total quantity of chemicals released from each facility was determined by calculating the total on-site releases for each individual chemical released by each facility. The facility totals within each cluster were then summed to obtain the total amount of chemicals released per cluster.

Statistical analysis

The clustering analyses, determination of the number of clusters, calculation of total chemical releases per facility and cluster, percentages for the SES variables by block group, and descriptive statistics for the SES variables by cluster were all done using R (Version 3.5.3) (R Core Team 2020). R was also used to geocode the location of the TRI facilities into the clusters, as well as create the map of the clusters with the facilities' locations highlighted. GraphPad Prism (Version 7.03) was used to make Fig. 2.

Results

Characterization of residential clusters

In 2000, Erie, Monroe, Onondaga, and Albany counties were comprised of 913, 601, 409, and 233 block groups, respectively, which totals 2,156 block groups across the four-county area. There were 33 block groups removed from the study because no population data was available; upon closer inspection, these block groups were located in areas containing businesses, churches, schools, or parks. This left 2,123 block groups to be included in this study.

Using the US Census variables selected (Table 2), seven unique residential clusters were formed (Fig. 3); others who used a similar clustering technique generated similar numbers of clusters in their work (Humphreys and Carr-Hill 1991; Mirowsky et al. 2017; Pedigo et al. 2011; Weaver et al. 2019). The demographic characteristics of each cluster can be seen in Fig. 2. For clarity, we named the clusters to reflect their predominant demographic characteristics. Our seven clusters include the following: Minority Working Class, Working Class, Wealthy Educated, Suburban, Low-SES Urban, Wealthy Working Class, and Rural. A detailed table of the cluster and countywide averages, separated by demographic category, can be seen in Online Resource 1 and 2, respectively.

Across the four counties, we observed similarities in how the newly formed clusters were distributed. Starting from the outskirts of each county and working inward (Fig. 3), the first cluster encountered is the Rural cluster. The Rural cluster has 189 block groups, and it is mostly rural with a comparatively higher percentage of owner-occupied housing and a comparatively lower percentage of households living below the poverty line. In moving closer to the more urban areas, the Wealthy Educated and Wealthy Working Class clusters are seen. The Wealthy Working Class ($n = 360$ block groups) and Wealthy Educated clusters ($n = 333$ block groups) have very similar characteristics, with the exception being the percent of individuals working non-managerial positions, in which the Wealthy Educated cluster had a lower percentage of those working in non-managerial positions than the Wealthy Working Class cluster. The next two cluster groups moving toward the urban centers are the Suburban and Working Class clusters. The Working Class cluster has 416 block groups, a comparatively lower percentage of individuals who were non-White, and a comparatively higher than average percentage of residents working in non-managerial positions. The Suburban cluster is comprised of 288 block groups, and it has a lower percentage of individuals living in owner-occupied housing, fewer single-parent households, and fewer who report working in non-managerial positions compared to the average found across the clusters. The last cluster surrounding the most urban area is the Minority Working Class cluster. The Minority Working Class cluster ($n = 411$ block groups) has a

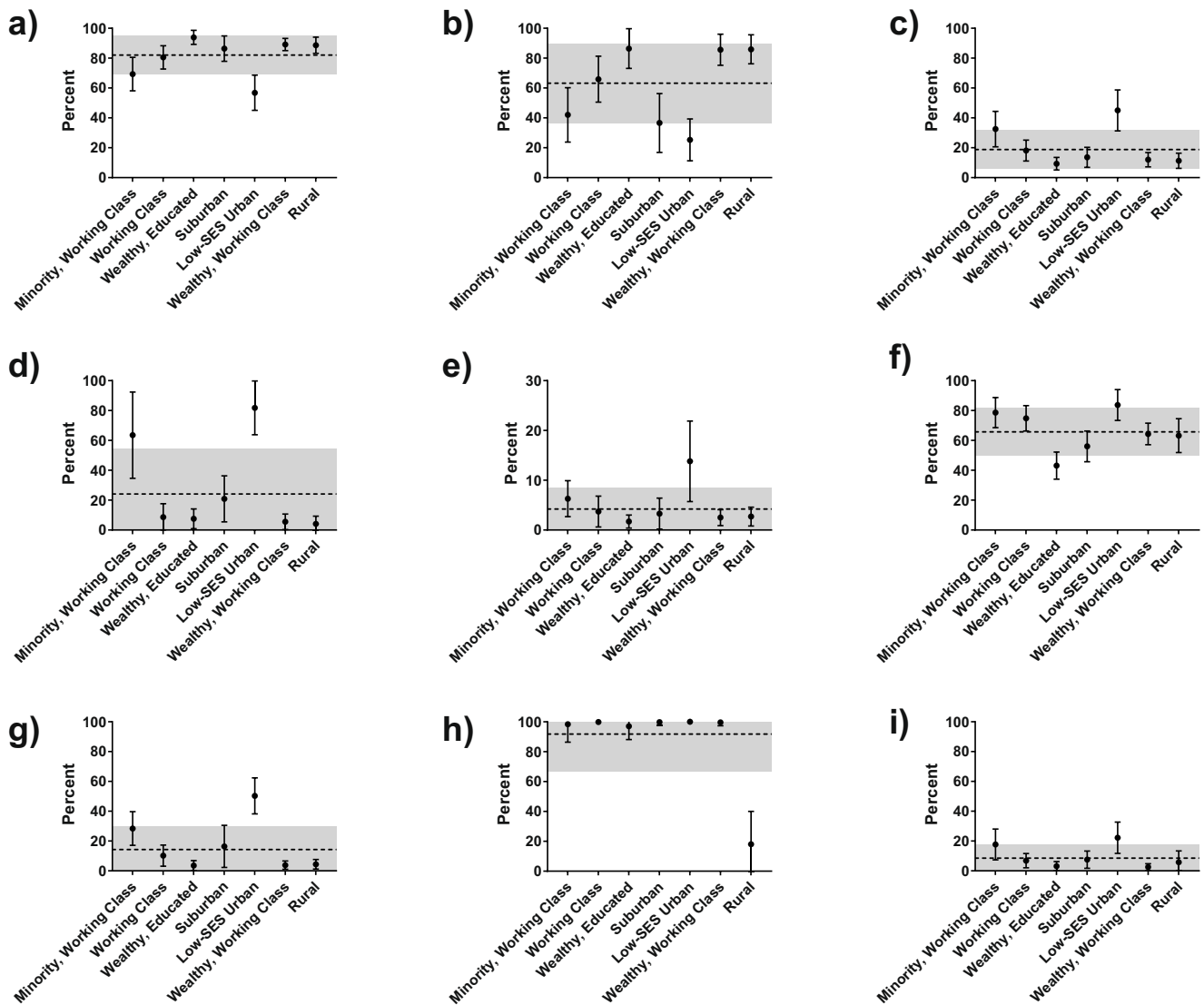


Fig. 2 Residential cluster information separated by the nine socioeconomic factors studied. Values represent mean \pm standard deviation. The dashed lines represent the average value for each factor, and the gray shaded areas represent the standard deviations from the

average values. **a** High school graduate. **b** Owner-occupied housing. **c** Single-parent housing. **d** Other race. **e** Unemployed. **f** Non-managerial positions. **g** Below poverty line. **h** Urban. **i** Vacant housing

comparatively higher than average percentage of individuals in non-managerial positions, has more non-White and unemployed individuals, has more households that live below the poverty line, and has more vacant housing. Lastly, the 126 block groups making up the Low-SES Urban cluster were 100% urban. This cluster had the lowest percentage of individuals in owner-occupied housing and high school graduates. This cluster also contained the highest percentage of single-parent households, unemployment, individuals working in non-managerial positions, households living below the poverty line, vacant housing, and those reporting as non-White.

The major Upstate New York city in each cluster is also identified in Fig. 3. In Albany, there were four TRI facilities.

This city was comprised of block groups that were mainly within the Minority Working Class cluster. In comparison, Buffalo, which is located in Erie County and has 12 TRI facilities, contains block groups that were from the Minority Working Class and Low-SES Urban clusters. In Monroe County, where the city of Rochester is located, combinations of block groups within the Minority Working Class, Low-SES Urban, and Suburban clusters can be found; only three TRI facilities were located in this city. A similar combination of block groups in the Minority Working Class, Low-SES Urban, and Suburban clusters is found in Syracuse (Onondaga County), which had only three TRI facilities located within it.

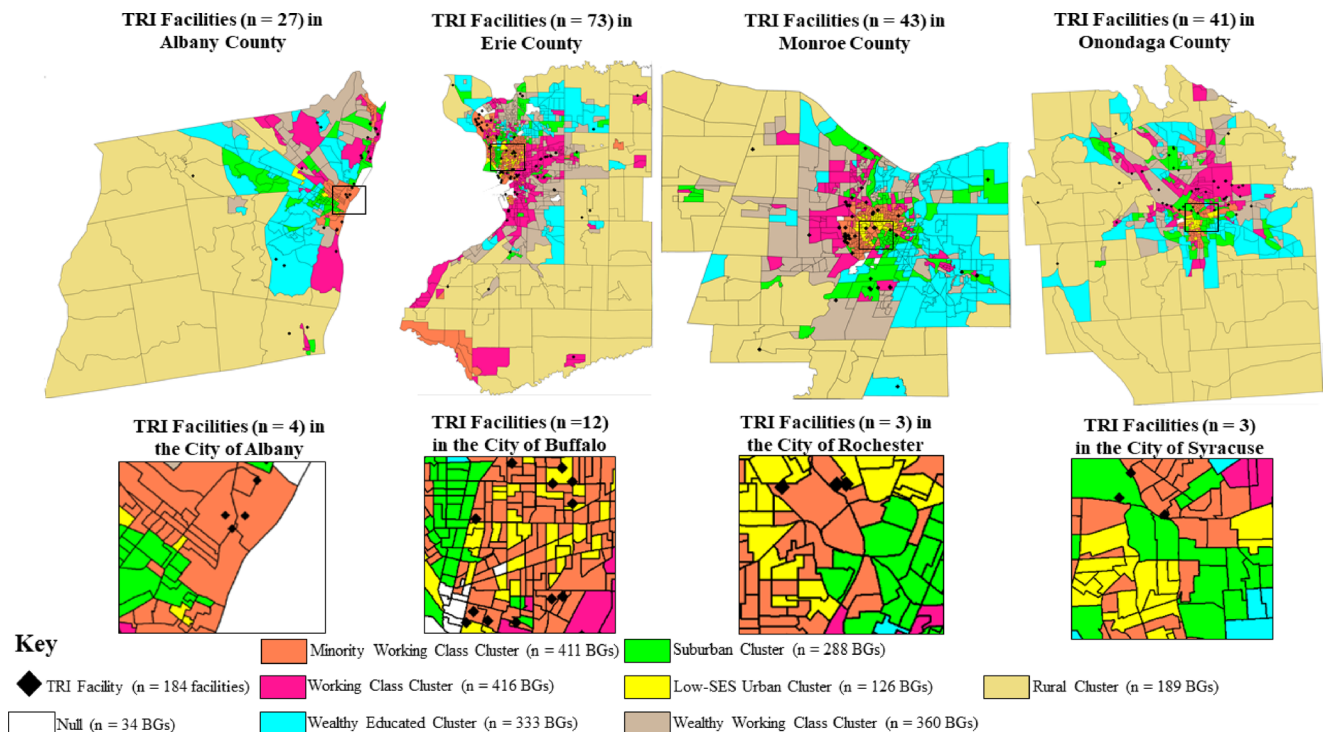


Fig. 3 Location of 184 Toxic Release Inventory (TRI) facilities within seven characteristically similar residential clusters comprised of 2123 Census block groups in Albany, Erie, Monroe, and Onondaga Counties.

Thirty-three block groups having partial or no population data were removed and are labeled as “Null”

Relationships between the residential clusters, TRI facility location, and pollutant releases

The location of each TRI facility was geocoded into its associated cluster (Fig. 3), and the number of polluting facilities per cluster was calculated (Table 3). Five TRI facilities were omitted because they were located in block groups that had no

Census demographic data, leaving 184 TRI facilities to be included in this analysis.

Overall, when looking at the number of facilities observed in each cluster, the clusters can be easily sub-divided into low ($n = 6$ and 8 TRI facilities), medium ($n = 16, 17,$ and 18 TRI facilities), and high ($n = 58$ and 60 TRI facilities) categories. When assessing for the lowest number of facilities present,

Table 3 Summary of cluster characteristics

	Minority Working Class	Working Class	Wealthy Educated	Suburban	Low-SES Urban	Wealthy Working Class	Rural	Total
Cluster area information								
Number of block groups (<i>n</i>)	411	416	333	288	126	360	189	2123
Land area (square miles)	81.12	195.01	355.75	109.07	12.24	282.85	1956.12	2992.16
Population information								
Population (<i>n</i>)	344,743	435,087	443,373	340,239	105,998	506,773	250,057	2,426,270
Population density (population per square mile)	4249.83	2231.07	1246.29	3119.43	8659.67	1791.70	127.83	810.88
TRI facility information								
Number of facilities (<i>n</i>)	58	60	6	19	8	17	16	184
Number of facilities per land area	0.71	0.31	0.02	0.17	0.65	0.06	0.01	0.06
On-site releases (pounds)	5,331,191	6,038,179	395,285	1,799,524	10,043	475,652	2,970,262	17,020,136
Amount of chemical emissions per land area (pounds per square mile)	65,719.81	30,963.43	1111.13	16,498.80	820.51	1681.64	1518.45	5688.24

these were found in the Wealthy Educated cluster ($n = 6$) and the cluster representing the Low-SES Urban demographic ($n = 8$ TRI facilities). These two clusters had no common characteristics apart from being urban. When looking at the clusters with a medium number of TRI facilities located within them, these clusters represent the Suburban ($n = 19$ TRI facilities) and Rural ($n = 16$ TRI facilities) areas, and the Wealthy Working Class ($n = 17$ TRI facilities) cluster. The Working Class ($n = 60$ TRI facilities) and Minority Working Class ($n = 58$ TRI facilities) clusters represented the clusters with the highest numbers of facilities and the greatest amount of chemical emissions per land area. Although the clusters are best described by the sum of their attributes rather than by the individual factors that make up each cluster (Mirowsky et al. 2017), these two clusters share the common trait of having a higher than average percentage of people working in non-managerial positions and being predominately located in urban areas (Fig. 2). The Minority Working Class cluster also had the highest number of facilities per area (0.71), although we do note that these are not distributed uniformly throughout the cluster.

Along with the number of facilities present in each cluster, the quantity of chemicals released (from the air, water, and on land) in each cluster was also examined and compared between the clusters (Table 3). The largest quantity of chemicals released was in the Working Class cluster, which also had the largest number of facilities. However, the smallest numbers of facilities were found in the Wealthy Educated cluster ($n = 6$), but this cluster did not release the smallest quantity of chemicals. Instead, the smallest quantity of chemicals, 10,000 pounds, was released in the Low-SES Urban cluster, which only had eight facilities present (Table 3). In Online Resource 3, the county maps with markers weighted by chemical releases can be observed.

Discussion

For this study, in four counties located in Upstate New York, seven unique clusters were created from nine commonly measured indicators of socioeconomic status using hierarchical clustering. The locations of the polluting facilities, and the quantities of pollutants they released, were geocoded into the clusters to assess which clusters had the largest and smallest number of facilities and chemical releases. We found that the greatest quantity of TRI facilities and on-site pollutant releases were found in two clusters representing working class communities. Interestingly, when looking for which clusters had the fewest number of facilities present, these were found in areas that were either wealthy and educated or of low socioeconomic status. Additionally, the lowest socioeconomic status cluster also had the lowest quantity of chemicals released within it, even though it did not have the lowest number

of facilities. Our results suggest there may be other SES indicators other than race and income that are associated with the location of polluting facilities, and that using just the number of TRI facilities in an area may not account for the quantity of potential exposures for the communities living around them.

Overall, we found that the greatest numbers of TRI facilities present in our study area were located in the Working Class and Minority Working Class clusters. Both these clusters share the common trait of having a higher than average percentage of people working in non-managerial positions (Fig. 2), giving rise to their names. It has been suggested by Ringquist (1997) and Wolverton (2009) that firms will often prefer a location that provides access to a large pool of inexpensive, available workers when siting for facilities, which could explain our findings. Additionally, our results support the idea that, when siting potential locations for facilities, the ease with which a plant can hire workers, and the qualifications of those workers, might also be factors they consider. Similar results to ours have been seen in other work (Kriesel et al. 1996; Ringquist 1997; Boer et al. 1997; Cutter et al. 2002) as well. However, we want to caution that the clusters are best described by the sum of their attributes rather than by their individual factors (Mirowsky et al. 2017), so we cannot say with certainty that it is only non-managerial positions that have influenced our finding. The only other factor that both of these clusters had in common was that they were both highly urban. In a study by Ringquist (1997), he found that exposure to TRI pollutants is especially high in urban areas with large numbers of manufacturing employees and low median income, but not in areas with high levels of poverty and large numbers of high school dropouts. Ringquist further explains that this paints a picture of these facilities being concentrated in urban, working class neighborhoods, which is in agreement with our results. TRI facilities being located within highly urban areas have also been documented by other research groups (Johnson et al. 2016; United States Environmental Protection Agency 2020b). Although we cannot with certainty determine a specific reason for our findings, it was clear that the two working class clusters overwhelmingly had the most facilities located within them compared to any of the other clusters that were created.

When assessing the clusters with the fewest number of facilities present, these were found in the Wealthy Educated cluster and the cluster representing the Low-SES Urban demographic. It was not surprising to see that the wealthiest cluster contained one of the lowest numbers of polluting facilities, as many studies have demonstrated that areas of higher SES are less likely to face environmental burdens than areas of lower SES (Wheeler and Ben-Shlomo 2005; Wilson et al. 2012; Johnson et al. 2016). However, having a low quantity of TRI facilities located in the most impoverished cluster (Low-SES urban) was not expected (Commission for Racial Justice 1987; Pastor Jr. et al. 2001; Mohai and Saha

2007; Bullard et al. 2007). There are several possibilities why this might have been observed. First, it is possible that lower numbers of facilities were found in the most deprived area because of the low racial diversity in Upstate New York compared to some of the other areas studied in the environmental justice literature. For example, in 2000, Upstate New York reported approximately 7% of the population as non-Hispanic Black and 2% reporting as non-Hispanic Asian (RLS Demographics 2011). In contrast, in the study done by Bullard et al. (2007), which found that minorities were disproportionately exposed to environmental hazards across the country, the area being studied was much more diverse with approximately 12% of the residents reporting as Black or African American and approximately 1% reporting as American Indian and Alaska Native (Grieco and Cassidy 2001). Additionally, a study by Pastor Jr. et al. (2001) was conducted in Los Angeles County, California, and for that work, which found that facilities were being sited in areas where minorities were already present, the reported population of Los Angeles County was 45% Hispanic (United States Census Bureau 2020). While these studies may have shown results that differ from ours, the racial diversity of their study areas did differ from that of Upstate New York, which could have influenced the findings.

It is also possible that having available and useable land to build facilities might factor into these results, since facilities can only be built where there is space to build them. In the current study, the Low-SES Urban cluster did have the lowest land area of all the clusters in this work (Table 3), with only 12 square miles encompassing 126 block groups. However, the Wealthy Educated cluster had the second highest land area (~356 square miles) of all the clusters created; the 333 block groups in the Wealthy Educated cluster ranged in size from 0.05 square miles to 12.20 square miles (United States Census Bureau 2019). Therefore, while it is possible that land area could make a difference in the placement of polluting facilities based on land availability, this might not be the only factor to explain our findings. For example, when calculating the number of facilities per square mile, the Low-SES Urban cluster had the second highest value, after the Minority Working Class cluster (Table 3), even though it had only eight TRI facilities located within it; the Rural cluster, which had the greatest land area, had only 0.01 facilities per square mile.

One of the main objectives of this study was to assess which neighborhoods were located near the highest and lowest quantity of chemicals released. In a paper by Ash and Fetter (2002), using the mass of pollutants reported from a facility was found to be a much better proxy of environmental equity than looking at just the presence of facilities in an area. This work was supported by research done by Collins et al. (2020), who demonstrated how a small number of facilities can be responsible for a large quantity of toxic releases within an industry category. In the current study, we found that the

largest quantity of chemicals released was from the Working Class cluster, which also had the largest number of facilities present within it. However, this was not the case when we looked at the clusters with the smallest number of facilities located within them, which were the Wealthy Educated ($n = 6$ TRI facilities) and Low-SES Urban ($n = 8$ TRI facilities) clusters. Although there is only a difference of two facilities between them, the six TRI facilities located within the Wealthy Educated cluster released approximately 40 times the quantity of chemicals compared to the eight facilities in the Low-SES Urban cluster. This result highlights how solely looking at the number of facilities rather than their releases might not be the best metric in terms of potential exposure to a population or neighborhood. Further, our result of small releases in our Low-SES Urban cluster is not a conclusion typically seen in other studies of a similar methodological framework. For example, in a study conducted by Sicotte (2010) in the Philadelphia Metropolitan Statistical Area, numerical points were assigned to 14 types of environmental hazards. Their findings suggest that only the most affluent escape proximity to hazards; however, this is in contrast to our results. While our most affluent cluster did release fewer chemicals than our two highest releasing clusters (Minority Working Class and Working Class), those living in the Wealthy Educated cluster could still be exposed to high-level environmental hazards.

Strengths and limitations

The present study has several limitations. First, the data used in this study was from 2000; this was because there was no block group demographic data available from the 2010 Census for all of the specific variables we were looking at. Next, the clusters created in this study are very specific to the variables we chose. Unfortunately, there is no consensus as to which variables best describe R-SES (Messer et al. 2006); however, many of the most commonly utilized variables from prior studies were included in our analyses. With respect to the TRI program, there are minimum reporting requirements, meaning that many smaller facilities are exempt from reporting their releases (Dolinoy and Miranda 2004; Wilson et al. 2012; United States Environmental Protection Agency 2019d). The TRI program also does not address the environmental fate or transport of industry emissions (Dolinoy and Miranda 2004). Direction of transport could be inferred based upon the wind direction or location of railways in the study location for the air releases, but that was beyond the scope of this study. Next, this study did not address the SES of the clusters at the time the facilities were being sited. Having this information could give us more information about the communities that existed before the facilities were built, and that could allow us to track, through time, how the neighborhoods have changed as a result of the facilities being present. This is a major limitation noted in several, similarly designed

research (Neumann et al. 1998; Dolinoy and Miranda 2004; Mohai et al. 2009; Wilson et al. 2012). One study by Pastor Jr. et al. (2001) stated that, over a 30-year period, toxic facilities tended to be located in vulnerable neighborhoods, not the other way around. Lastly, using the unit-hazard coincidence method has limitations associated with it, such as assuming that the potential hazards of living close to a TRI facility are uniform across the entire block group. However, we have chosen to use block groups compared to Census tracts to alleviate some of those concerns.

There are several strengths from this study that should be recognized. First, this study used hierarchical clustering to form the geographical neighborhoods. Clustering is a more novel technique to examine R-SES, but it has been successfully used in a handful of studies (Humphreys and Carr-Hill 1991; Mirowsky et al. 2017; Pedigo et al. 2011; Roussot et al. 2016; Weaver et al. 2019). Compared to other studies that use principal component analysis (PCA) (Messer et al. 2010; Pampalon et al. 2012; Berkowitz et al. 2015), clustering our data lets us utilize all the demographic indicators we assume to be important in our analysis, rather than mathematically reducing that number of indicators using statistical techniques. The current study also utilized US Census block groups to create the clusters. Some studies have been performed to determine the appropriate geographic scale for area-level SES disparities and have found scaling to be one of the most important factors in this type of analysis (Perlin et al. 1995; Dolinoy and Miranda 2004). Block groups are a smaller geographic area than census tracts; there are more studies assessing R-SES that are done at the census tract level (Reagan and Salsberry 2005; Wilson et al. 2012; Berkowitz et al. 2015) than at the block group level (Berkowitz et al. 2015; Cutter et al. 1996; Dolinoy and Miranda 2004; Mirowsky et al. 2017; Weaver et al. 2019). Finally, our study looked at both the location of these TRI facilities and the quantity of their emissions. To examine the burden of TRI facilities on a community, basic emission levels were incorporated into our work, allowing for both an in-depth analysis of not just the location of these facilities, but also the amount of chemicals they are releasing into their surrounding areas.

Conclusions and future work

In conclusion, we completed a multi-disciplinary environmental justice study to identify seven unique residential clusters in Albany, Erie, Monroe, and Onondaga Counties based on nine US Census demographic variables using hierarchical clustering. This clustering technique has been used in very few studies and allowed us to compare a large geographic area on a much smaller scale. It also allowed us to include all variables we deemed influential in our study rather than mathematically

reducing them. Using TRI data, we were able to identify all facilities within these four counties and show a relationship between the residential cluster characteristics and the location of these TRI facilities. The population that seemed to be most impacted by both the location of polluting facilities and facility releases was the working class, which, in this current study, represents a community that had a comparatively lower percentage of individuals who were non-White, and a comparatively higher-than-average percentage of residents working in non-managerial positions. Our most impoverished cluster had one of the lowest number of polluting facilities located within it and had the lowest number of releases present, suggesting that, in addition to race and class as predictors of community-level contamination, other metrics of socioeconomic status might be influential when assessing the complex landscape of environmental inequity from different geographic areas. In addition, our work also highlights that it might be a combination of factors that make up a community, rather than just one identifiable factor such as race or income, that should be considered in future environmental justice work of this nature. Future work on this project includes a more in-depth look at the chemicals being released from these facilities—including their release type (i.e., air, water, land) and toxicity—to determine the potential risk of exposure among these communities.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13412-021-00664-7>.

Acknowledgments We would like to acknowledge the Discovery Challenge Seed Grant and the Center for Environmental Medicine and Informatics (CEMI) for helping fund this project. Also, we would like to acknowledge Dustin Hill for his involvement and help during this project.

Author contribution Amanda Charette: methodology; formal analysis and investigation; writing—original draft preparation; writing—review and editing

Mary Collins: writing—review and editing; funding acquisition

Jaime Mirowsky: conceptualization; methodology; writing—review and editing; funding acquisition; supervision

Funding This work was supported by the Center for Environmental Medicine and Informatics (CEMI) at the State University of New York College of Environmental Science and Forestry.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Code availability All analyses were done using RStudio version 3.5.3. All code used in creating images is available from the corresponding author upon request.

References

- Arora S, Cason TN (1999) Do community characteristics influence environmental outcomes? Evidence from the Toxics Release Inventory. *South Econ J* 65:691–716. <https://doi.org/10.2307/1061271>
- Ash M, Fetter TR (2002) Who lives on the wrong side of the environmental tracks? Evidence from the EPA's Risk-Screening Environmental Indicators Model. Amherst, MA
- Berkowitz SA, Traore CY, Singer DE, Atlas SJ (2015) Evaluating area-based socioeconomic status indicators for monitoring disparities within health care systems: results from a primary care network. *Health Serv Res* 50:398–417. <https://doi.org/10.1111/1475-6773.12229>
- Boer JT, Pastor M, Sadd JL, Snyder LD (1997) Is there environmental racism? The demographics of hazardous waste in Los Angeles County. *Soc Sci Q (University Texas Press)* 78:793–810
- Bolin B, Nelson A, Hackett EJ, Pijawka KD, Smith CS, Sicotte D, Sadalla EK, Matranga E, O'Donnell M (2002) The ecology of technological risk in a Sunbelt City Identity Symbolism View project History of Environmental Inequality in the Philadelphia Region View project. *Environ Plan A* 34:317–339. <https://doi.org/10.1068/a3494>
- Bullard RD, Mohai P, Saha R, Wright B (2007) Toxic wastes and race at twenty: 1987 - 2007. *Environmental Law* 38(2), 371–411
- Collins M, Pulver S, Hill D, Manski B (2020) Characterizing disproportionality in facility-level toxic releases in US manufacturing, 1998–2012. *Environ Res Lett* 15. <https://doi.org/10.1088/1748-9326/ab7393>
- Commission for Racial Justice (1987) Toxic wastes and race in the United States. <https://www.nrc.gov/docs/ML1310/ML13109A339.pdf>
- Curry SJ, Wagner EH, Cheadle A et al (1993) Assessment of community-level influences on individual's attitudes about cigarette smoking, alcohol use, and consumption of dietary fat. *Am J Prev Med* 9:78–84
- Cutter SL, Holm D, Clark L (1996) The role of geographic scale in monitoring environmental justice. *Risk Anal* 16:517–526. <https://doi.org/10.4324/9781849771542>
- Cutter SL, Scott MS, Hill AA (2002) Spatial variability in toxicity indicators used to rank chemical risks. *Am J Public Health* 92:420–422. <https://doi.org/10.2105/AJPH.92.3.420>
- Dolinoy DC, Miranda ML (2004) GIS modeling of air toxics releases from TRI-reporting and non-TRI-reporting facilities: impacts for environmental justice. *Environ Health Perspect* 112:1717–1724. <https://doi.org/10.1289/ehp.7066>
- Evans GW, Marcynyszyn LA (2004) Environmental justice, cumulative environmental risk, and health among low- and middle-income children in Upstate New York. *Am J Public Health* 94:1942–1944. <https://doi.org/10.2105/AJPH.94.11.1942>
- Fricker R Jr, Hengartner NW (2001) Environmental equity and the distribution of toxic release inventory and other environmentally undesirable sites in metropolitan New York City. *Environ Ecol Stat* 8: 33–52. <https://doi.org/10.1023/A:1009649815643>
- Friedman HP, Rubin J (1967) On some invariant criteria for grouping data. *J Am Stat Assoc* 62:1159–1178. <https://doi.org/10.1080/01621459.1967.10500923>
- Grieco EM, Cassidy RC (2001) Overview of race and Hispanic origin. US Dept of Commerce, Economics and Statistics Administration. <https://www.census.gov/prod/2001pubs/c2kbr01-1.pdf>
- Hill DT, Collins MB, Vidon ES (2018) The environment and environmental justice: linking the biophysical and the social using watershed boundaries. *Appl Geogr* 95:54–60. <https://doi.org/10.1016/j.apgeog.2018.04.007>
- Humphreys K, Carr-Hill R (1991) Area variations in health outcomes: artefact or ecology. *Int J Epidemiol* 20:251–258
- Johnson R, Ramsey-White K, Fuller CH (2016) Socio-demographic differences in toxic release inventory siting and emissions in Metro Atlanta. *Int J Environ Res Public Health* 13:1–12
- Kriesel W, Centner TJ, Keeler AG (1996) Neighborhood exposure to toxic releases: are there racial inequities? *Growth Change* 27:479–499
- Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, Elo I, Burke JG, O'Campo P (2006) The development of a standardized neighborhood deprivation index. *J Urban Heal* 83:1041–1062. <https://doi.org/10.1007/s11524-006-9094-x>
- Messer LC, Oakes JM, Mason S (2010) Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. *Am J Epidemiol* 171:664–673. <https://doi.org/10.1093/aje/kwp435>
- Mirowsky JE, Devlin RB, Diaz-Sanchez D, Cascio W, Grabich SC, Haynes C, Blach C, Hauser ER, Shah S, Kraus W, Olden K, Neas L (2017) A novel approach for measuring residential socioeconomic factors associated with cardiovascular and metabolic health. *J Expo Sci Environ Epidemiol* 27:281–289. <https://doi.org/10.1038/jes.2016.53>
- Mohai P, Saha R (2007) Racial inequality in the distribution of hazardous waste: a national-level reassessment. *Soc Probl* 54:343–370. <https://doi.org/10.1525/sp.2007.54.3.343>
- Mohai P, Pellow D, Roberts JT (2009) Environmental justice. *Annu Rev Environ Resour* 34:405–430. <https://doi.org/10.1146/annurev-environ-082508-094348>
- Neumann CM, Forman DL, Rothlein JE (1998) Hazard screening of chemical releases and environmental equity analysis of populations proximate to toxic release inventory facilities in Oregon. *Environ Health Perspect* 106:217–226. <https://doi.org/10.1289/ehp.98106217>
- Pampalon R, Hamel D, Gamache P et al (2012) An area-based material and social deprivation index for public health in Québec and Canada. *Can J Public Heal* 103:S17–S22. <https://doi.org/10.17269/cjph.103.3156>
- Pastor M Jr, Sadd J, Hipp J (2001) Which came first? Toxic facilities, minority move-in, and environmental justice. *J Urban Aff* 23:1–21. <https://doi.org/10.1111/0735-2166.00072>
- Pedigo A, Seaver W, Odoi A (2011) Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches. *PLoS One* 6: e22693. <https://doi.org/10.1371/journal.pone.0022693>
- Perlin SA, Setzer RW, Creason J, Sexton K (1995) Distribution of industrial air emissions by income and race in the United States: an approach using the Toxic Release Inventory. *Environ Sci Technol* 29: 69–80. <https://doi.org/10.1021/es00001a008>
- R Core Team (2020) R: A language and environment for statistical computing. Vienna, Austria. URL: <https://www.R-project.org/>
- Reagan PB, Salsberry PJ (2005) Race and ethnic differences in determinants of preterm birth in the USA: broadening the social context. *Soc Sci Med* 60:2217–2228
- Ringquist EJ (1997) Equity and the distribution of environmental risk: the case of TRI facilities. *Soc Sci Q* 78:811–829
- RLS Demographics I (2011) Population distribution and diversity in New York State. <https://www.empirecenter.org/wp-content/uploads/2014/11/PopChange-2000-2010.pdf>. Accessed 20 Jul 2020
- Roberts EM (1997) Neighborhood social environments and the distribution of low birthweight in Chicago. *Am J Public Health* 87:597–603. <https://doi.org/10.2105/AJPH.87.4.597>
- Roussot A, Cottenet J, Gadreau M, Giroud M, Béjot Y, Quantin C (2016) The use of national administrative data to describe the spatial distribution of in-hospital mortality following stroke in France, 2008–2011. *Int J Health Geogr* 15. <https://doi.org/10.1186/s12942-015-0028-2>
- Sicotte D (2010) Some more polluted than others: unequal cumulative industrial hazard burdens in the Philadelphia MSA, USA. *Local*

- Environ 15:761–774. <https://doi.org/10.1080/13549839.2010.509384>
- United States Census Bureau (2019) American FactFinder. <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>. Accessed 10 Jun 2019
- United States Census Bureau (2020) Racial/ethnic population by census years for Los Angeles County, California. In: Los Angeles Alm. <http://www.laalmanac.com/population/po13.php>. Accessed 20 Jul 2020
- United States Environmental Protection Agency (2019a) What is EPCRA? Emergency Planning and Community Right-to-Know Act (EPCRA). <https://www.epa.gov/epcra/what-epcra>. Accessed 5 Nov 2019
- United States Environmental Protection Agency (2019b) What is the Toxic Release Inventory. <https://www.epa.gov/toxics-release-inventory-tri-program/what-toxics-release-inventory>. Accessed 5 Nov 2019
- United States Environmental Protection Agency (2019c) Environmental justice. <https://www.epa.gov/environmentaljustice>. Accessed 5 Jan 2020
- United States Environmental Protection Agency (2019d) Reporting for TRI Facilities. <https://www.epa.gov/toxics-release-inventory-tri-program/reporting-tri-facilities>. Accessed 7 Oct 2019
- United States Environmental Protection Agency (2020a) Release chemical report. https://enviro.epa.gov/triexplorer/tri_release.chemical. Accessed 10 Jun 2020
- United States Environmental Protection Agency (2020b) States and Metropolitan Areas | Toxics Release Inventory (TRI) National Analysis | US EPA. <https://www.epa.gov/trinationalanalysis/states-and-metropolitan-areas>. Accessed 17 Nov 2020
- Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Weaver AM, McGuinn L, Neas L et al (2019) Neighborhood sociodemographic effects on the associations between long-term PM2.5 exposure and cardiovascular outcomes and diabetes mellitus. *Environ Epidemiol* 3:e038. <https://doi.org/10.1097/ee9.000000000000038>
- Wheeler BW, Ben-Shlomo Y (2005) Environmental equity, air quality, socioeconomic status, and respiratory health: a linkage analysis of routine data from the Health Survey for England. *J Epidemiol Community Heal* 59:948–954. <https://doi.org/10.1136/jech.2005.036418>
- Williams MM (2008) Linking health hazards and environmental justice: A case study in Houston, Texas. Dissertation, University of South Florida
- Wilson SM, Fraser-Rahim H, Williams E, Zhang H, Rice LS, Svendsen E, Abara W (2012) Assessment of the distribution of toxic release inventory facilities in Metropolitan Charleston: an environmental justice case study. *Am J Public Health* 102:1974–1980. <https://doi.org/10.2105/AJPH.2012.300700>
- Wolverton A (2009) Effects of Socio-Economic and Input-Related Factors on Polluting Plants' Location Decisions. *BE J Econ Anal Policy* 9:9. <https://doi.org/10.2202/1935-1682.2083>
- World Population Review (2020) Poorest cities in America. In: *World Popul. Rev.* <https://worldpopulationreview.com/us-cities/poorest-cities-in-america/>. Accessed 18 Jun 2020

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.