Published in final edited form as:
Am J Cardiol. 2023 August 15; 201: 150–157. doi:10.1016/j.amjcard.2023.06.037.

Geographical Patterns and Risk Factor Association of Cardio-Oncology Mortality in the United States

Issam Motairek, MD[a], Weichuan Dong, PhD[b], Pedro RVO Salerno, MD[a], Scott E. Janus, MD[a], Sarju Ganatra, MD[c], Zhuo Chen, PhD[a], Avirup Guha, MD[d], Mohamed He Makhlouf, MD[a], Neda Shafiabadi Hassani, MD[a], Sanjay Rajagopalan, MD[a,e], Sadeer G. Al-Kindi, MD[a,e,*]

[a]Harrington Heart and Vascular Institute, University Hospitals Cleveland Medical Center, Cleveland, Ohio

[b]Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio

[c]Cardio-Oncology Program, Lahey Clinic, Burlington, Massachusetts

[d]Cardio-Oncology Program, Georgia Cancer Center, Medical College of Georgia at Augusta University, Augusta, Georgia

[e]Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, Ohio

Abstract

Cardio-oncology mortality (COM) is a complex issue that is compounded by multiple factors that transcend a depth of socioeconomic, demographic, and environmental exposures. Although metrics and indexes of vulnerability have been associated with COM, advanced methods are required to account for the intricate intertwining of associations. This cross-sectional study utilized a novel approach that combined machine learning and epidemiology to identify high-risk sociodemographic and environmental factors linked to COM in United States counties. The study consisted of 987,009 decedents from 2,717 counties, and the Classification and Regression Trees model identified 9 county socio-environmental clusters that were closely associated with COM, with a 64.1% relative increase across the spectrum. The most important variables that emerged from this study were teen birth, pre-1960 housing (lead paint indicator), area deprivation index, median household income, number of hospitals, and exposure to particulate matter air pollution. In conclusion, this study provides novel insights into the socio-environmental drivers of COM

[*]Corresponding author: Tel: 216-844-1000; fax: 216-844-8081. Sadeer.Al-Kindi@uhhospitals.org (S.G. Al-Kindi).

This work was partly funded by the National Institute on Minority Health and Health Disparities, Bethesda, Maryland, award number P50MD017351.

See page 156 for Declaration of Conflict of Interest.

Declaration of Competing Interest
Dr. Dong is supported by contracts from the Cleveland Clinic Foundation, including a subcontract from Celgene Corporation. The remaining authors have no conflicts of interest to declare.

Supplementary materials
Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.amjcard.2023.06.037.

and highlights the importance of utilizing machine learning approaches to identify high-risk populations and inform targeted interventions for reducing disparities in COM.

In 2020 alone, 659,000 people died from cardiovascular disease (CVD) and 602,350 people died from cancer in the United States (US).[1] At the intersection of the top 2 causes of mortality in the country, lies cardio-oncology. Cardio-oncology is a rapidly expanding field because of the number of shared risk factors between CVD and oncology,[2] the known cardiotoxicity of cancer therapies,[3] and the growing number of cancer survivors.[4] Cardiovascular toxicity from cancer therapies, besides impacting the physical and psychosocial health status of patients, can sometimes pose a greater risk of death than the specific cancer of the patient.[5] Current guidelines suggest a 3-step approach to the dynamics of cardiovascular toxicity risk in patients with cancer: baseline risk evaluation, cancer treatment surveillance, and long-term follow-up after cancer treatment. However, adequate implementation and access to optimal follow-up is still a challenge.[5] Disparities in sociodemographic and environmental determinants of health (SEDH) can help explain some of the disparities in cardio-oncology mortality (COM). Previous small studies have explored a limited number of possible factors and indexes.[6,7] However, because of the complex interactions of SEDH with COM and limitations of conventional statistical methods, previous studies have yet to consider a large number of risk factors covering multiple domains of SEDH that could potentially be linked to COM. A comprehensive appreciation of the entire spectrum of SEDH would allow construction and adaptation of healthcare policies to target the essential roots of disparities in COM and explore high-risk sociodemographic and environmental risk factors associated with COM across US counties. Using advanced machine learning models, we intended to explore associations between county-level SEDH and COM to disentangle their complex intersections and provide analytical frameworks for future studies on risk factors of COM.

Methods

We extracted a wide range of county-level sociodemographic and environmental exposures from multiple sources. Machine learning approaches were used to identify combinations of characteristics highly associated with COM, which were defined as county clusters of COM. Geographic information systems were used to map these clusters and to identify areas with least favorable outcomes, offering opportunities for targeted public health interventions and resource allocation.

We examined COM by utilizing the publicly available multiple causes of death files maintained by the National Center for Health Statistics through the CDC-WONDER (Centers for Disease Control and Prevention Wide-ranging Online Data for Epidemiologic Research) database. CDC-WONDER maintains mortality data based on death certificate information for all 50 states, categorizing the cause of death using the International Classification of Disease, 10th version (ICD-10). In the multiple causes of death category, records were included if they presented with any of the ICD-10 codes for CVD mortality (ischemic heart disease [I20-I25], heart failure [I50], cerebrovascular diseases [I60-I69], and hypertensive heart disease [I10-I15]) and an ICD-10 code for cancer mortality [malignant neoplasms (C00-C97)]).

Our population was restricted to decedents aged 15 years and older who died from CVD and cancer between 2016 and 2020. County-level age-adjusted COM was calculated as deaths per 100,000 individuals standardized to the 2000 US Standard Population. Counties or equivalents from Hawaii and Alaska were excluded because of the scarcity of multiple social and environmental variables.

A total of 71 key SEDH variables potentially associated with COM were harvested from previously published sources[8,9] (Table 1). Data from 2017 best harmonized sociodemographic and environmental data and were utilized in our study.

The environmental indicators were collected from the EPA-EJSCREEN (Environmental Protection Agency-Environmental Justice Screening tool). There were 11 environmental indicators in the 2020 version of EJSCREEN covering different time points (2014 to 2020). The indicators represent census block group level exposures, and thus county-level exposures were estimated by applying the advised method by the technical documentation guide of EJSCREEN.

The 56 sociodemographic variables used were gathered from the Area Health Resources Files and County Health Rankings & Roadmaps. The harvested variables span diverse fields including access to healthcare, behavioral risk factors, population characteristics, and other health-related variables. Furthermore, we obtained county-level area deprivation index (ADI), a measure of neighborhood deprivation and social vulnerability incorporating 17 census variables.[10] We additionally adopted 3 subcategories of area deprivation themes: financial strength, economic hardship and inequality, and educational attainment (para 3, Berg et al[11]).

We utilized Classification and Regression Tree (CART) to identify county clusters or combinations of characteristics most associated with COM rates at the county level. We then used random forest (RF) analysis to evaluate the relative importance of variables in predicting COM and to determine whether the most important variables were captured by CART.

CART is a machine learning model that sequentially divides data into smaller and more homogenous groups using binary conditional inferences (“if-then” rules) to predict certain outcomes.[8] Pearson’s correlation is used at each branch point to check for statistical significance. When certain thresholds (stopping criterion) are reached, CART stops partitioning data, and clusters of homogenous counties are formed that satisfy a collection of conditional inferences. In our study, we set the following CART thresholds: maximum tree depth of 6 splits, a minimum number of 200 counties in each terminal node, and a statistical significance (α <0.05) for each branching point. We also conducted a sensitivity analysis for our approach, using a smaller minimum number of terminal node counties (100) while utilizing the same approach.

The terminal nodes (leaves) consist of clusters of counties that meet the same inferences and have similar COM. The unique combination of characteristics along the path from the top split to a terminal node, determined by the conditional inferences met, represents a COM cluster of the counties in that node. These clusters were then labeled using alphabet letters from left to right of the tree. Our model was validated against a random 20% hold-out sample comparing the COM rates between training and testing using box plots.

Similar to CART, RF uses recursive partitioning. Instead of relying on one tree, it creates and aggregates multiple trees using random variable selection and bootstrap sampling. It takes the average of the outputs of these trees as a prediction and calculates the relative importance of variables in the detection of COM according to the mean decrease in node impurity. In total 20,000 trees were created incorporating 71 variables, and the number of randomly sampled variables at each tree split was set to 5.

We finally plotted the identified county clusters and COM to allow proper understanding of the cluster's geographic distribution across the US.

Statistical and machine learning analyses were made using open-access R software version 4.1.2 (The R Foundation for Statistical Computing, Vienna, Austria), and QGIS v 3.22.3. (QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation. URL http://qgis.org). A p <0.05 was considered statistically significant. No individual-level data were used, and thus institutional review board approval was not required.

Results

The study included a total of 987,009 decedents from 2,717 US counties who died from CVD and cancer. CART analysis, through a training set of 2,175 counties, identified 9 terminal nodes or clusters that share similar mortality and SEDH characteristics (Figure 1). These clusters were labeled with alphabet letters (A to I) by increasing median age-adjusted COM rates with a 64.1% relative increase in COM across the clusters (from 52.7 to 86.5 per 100,000 individuals, comparing clusters A and I). From 71 SEDH evaluated (listed in Table 1 along with their sources), the algorithm selected 6 key variables to serve as 8 splitting branches in our tree (*Teen birth, Pre-1960 Housing, Area deprivation index [ADI], Median Household Income, Hospitals, and particulate matter [PM]*$_{2.5}$ *exposure*). Supplementary Table 1 lists detailed information on each variable used in our study, including its source, description, year, and baseline mean value for the counties.

Figure 1 represents the findings from CART and Table 2 lists the summary characteristics of the county clusters. *Teen Birth* was used as the first node by CART and was the most important SEDH risk factor for COM, dividing the tree into the right side (clusters E, F, H, I) and the left side (clusters A, B, C, G, D).

On the right side of the tree (*Teen Birth* rates>26.9 per 100,000) The following nodes differed based on a further stratification of *Teen Birth* rates. Indeed, the sole use of *Teen Birth* rates >41.3 per 100,000 was sufficient to identify the cluster with the highest COM rates (cluster I – 86.5 per 100,000 individuals). Furthermore, when *Teen Birth* rates were between 26.9 and 41.3 per 100,000, CART utilized the number of hospitals per 100,000 individuals and the exposure to $PM_{2.5}$ to identify the clusters E, F, and H (fifth, fourth, and second highest COM rates, respectively).

On the left side of the tree (*Teen Birth* rates ≤26.9 per 100,000), CART additionally used Pre-1960 Housing, ADI, and Median Household Income to identify clusters A, B, C, G, and D. Which had the ninth, eighth, seventh, third, and sixth highest mortality rates, respectively. The combination of *Teen Birth* rates ≤26.9 per 100,000, ADI ≤90.3, and Pre-1960 Housing ≤18.3% (cluster A) or between 18.3% to 36.3% (cluster B) and could identify the lowest and 2nd lowest COM rates, respectively.

Next, the same regression tree was applied to the validation set, where the constituent counties fell into the same 9 categories as the derivation set. The comparisons in COM rates between the validation and derivation sets showed no significant differences across the 9 clusters (Supplementary Figure 1).

CART sensitivity analysis with a minimum number of 100 counties had more splitting and terminal nodes in the output (Supplementary Figure 2), showing a trend of a more complex relation between the SEDH variables to predict county-level COM. The additional variables include physically unhealthy days and access to exercise. The increased complexity of the mode lead to a 76.8% difference between the highest and lowest COM groups (median 92.4 vs 51.3). Supplementary Figure 3 shows the comparisons between training and test datasets to depict the performance of the sensitivity analysis model.

The geographic distributions of the county-level COM and their clusters are showcased in Figure 2. We observed that the higher county-level COM rates were mostly located in the Southern states. A lot of these counties correspond to high-risk clusters H and I.

Figure 3 highlights the importance of each SEDH variable with respect to COM mortality and indicates that variables generated by CART are top performers in the RF model. *Teen births (*1st*), Pre-1960 Housing* (4th), *Median Household Income* (6th), and *Hospitals* (7th). Other important variables include financial strength, population size, and lack of insurance for those aged between 18 and 64 years.

Discussion

In this analysis, we utilized machine learning approaches to unravel the socioeconomic and environmental determinants of health associated with COM. Using CART analysis, we uncovered SEDH clusters according to COM risk consisting of teen birth, pre-1960 housing, ADI, median household income, hospitals, and PM of size ≤2.5 μm. The same variables were demonstrated to be important variables in the RF analysis. Thereby demonstrating the ability of machine learning to help identify associations and improve understanding of the complex social and environmental factors associated with geographical disparities in disease burden.

As the field of cardio-oncology continues to rapidly expand, addressing COM and developing new tools to identify risk factors is imperative. With 5-year survival rates of children and adolescents with cancer exceeding 80%, long-term health effects are of great importance.[12] Understanding and recognizing those at greater risk has become a major concern and risk stratification classifications have been proposed, however, most take into consideration primarily clinical data.[5] Recent literature has demonstrated significant disparities in COM mostly based on race and ethnicity.[6,13,14] A limited number of investigations in sociodemographic determinants of COM have been undertaken, and found variables such as population density, lower income, and illiteracy are generally associated with CVD outcomes.[15–17] Unfortunately, even fewer have targeted cardio-oncology exclusively, likely because of the number of complicated and multidisciplinary factors including the intricate interplay between SEDH and COM.

Environmental exposures, specifically air pollutants, are associated with poor health outcomes.[18,19] Air pollution is estimated to cause 9 million annual deaths across the globe.[20] Despite the robust evidence linking environmental exposures with disease burden, environmental factors are often left out of disease prediction models. We incorporated environmental exposures with social determinants of health to build a more robust socio-environmental model for predicting COM. We demonstrated that in the 5 variables our CART analysis generated, 2 variables were environmental including pre-1960 housing (marker of lead exposure) and $PM_{2.5}$ (air pollutant). Pre-1960 housing is not only associated with lead exposure toxicity,[21,22] but can also act as a representative of poor socioeconomic status and potential environmental exposure hazards.[23] Further, $PM_{2.5}$ has been associated with increased mortality in patients with both CVD and cancer by means of various mechanisms including oxidative stress and systemic inflammation.[24,25]

We additionally show that teen births, ADI, and median household income are associated with COM. These markers are not necessarily causative agents of COM but may rather represent surrogates for poor social vulnerability. For example, teen births, have been previously associated with lower income, unemployment, and less educational attainment,[26] all of which may predispose individuals to increased risks of cardiovascular mortality.[6] Our findings suggest that certain social determinants of health may have a greater impact on COM than common risk factors like smoking, obesity, and physical inactivity. This could be because information regarding these risk factors is often embedded within social determinants of health. For instance, socially vulnerable individuals may be more likely to engage in unhealthy behaviors and adopt sedentary lifestyles, leading to higher rates of obesity and smoking. In addition, the number of hospitals per 100,000 was positively associated with COM, which potentially can be explained by the higher detection rates of cancer and CVD in areas with higher hospital densities and be confounded by other factors such as population density and the associated environmental exposures in densely populated areas.

The underlying causes of death by CVD and cancer are complex and are often intertwined with numerous sociodemographic, behavioral, and environmental factors. Incorporating a vast number of health-related variables from multiple sources is an advantage of tree-based machine learning models over traditional statistical methods (such as logistical regression) because they are not affected by potential correlations in independent variables (i.e., multi-collinearity). Previous studies have demonstrated the power of these tree-based machine learning approaches in uncovering clusters of late-stage cancer diagnosis[27] and premature cardiovascular mortality.[8] Our study further demonstrated the utility of CART and RF in exploring the associations between COM and sociodemographic and environmental risk factors. Given that the clusters identified in this study tend to be clustered in space (i.e., high-risk clusters are prevalent in the South, and low-risk clusters are prevalent in the Northeast), risk factors and COM may have place-dependent associations. Future studies should look at risk factor associations by US region and use the geographic RF model to explore these place-dependent associations as demonstrated in a previous study on cancer mortality.[9]

There are several limitations to this study. First, there might be inaccuracies in the cause of death identified by the ICD-10 codes found in the death certificates. Second, because of the nature of this cross-sectional study, casualty cannot be established between risk factors and COM. However, establishing association is still the first step for epidemiological studies to examine risk factors before causal relations can be established. Moreover, because of the absence of data, other common risk factors for COM were not included. Nevertheless, incorporating nontraditional risk factors provide valuable insights into the complex factors contributing to COM mortality. Furthermore, the data collection periods for SEDH variables and COM are not temporally consistent. Also, there might be latent effect of SEDH variables on COM that cannot be captured by the present study. We acknowledge this limitation and minimize it using a small range of years (2016 to 2020) for COM and closest years for the explanatory variables. Because of this, some counties were excluded because of the small number of cases per data user agreement for patient privacy concerns. Finally, the use of county-level data does not consider within-county variations of COM and risk factors, especially in large metropolitan counties. Future studies should investigate the association between COM and risk factors on a small geographic scale.

In conclusions, sociodemographic and environmental exposures have a complex relation with COM, and machine learning approaches can deconstruct this relation and demonstrate associations to allow improved understanding of the socio-environmental drivers of COM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. CDC. An update on cancer deaths in the United States Available at: https://www.cdc.gov/cancer/dcpc/research/update-on-cancer-deaths/index.htm. Accessed on September 20, 2022.
2. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared risk factors in cardiovascular disease and cancer. Circulation 2016;133:1104–1114. [PubMed: 26976915]
3. Rothe D, Paterson I, Cox-Kennett N, Gyenes G, Pituskin E. Prevention of cardiovascular disease among cancer survivors: the role of pre-existing risk factors and cancer treatments. Curr Epidemiol Rep 2017;4:239–247.
4. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, Jemal A, Kramer JL, Siegel RL. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019;69:363–385. [PubMed: 31184787]

5. Lyon AR, López-Fernández T, Couch LS, Asteggiano R, Aznar MC, Bergler-Klein J, Boriani G, Cardinale D, Cordoba R, Cosyns B, Cutter DJ, Azambuja E de, Boer RA de, Dent SF, Farmakis D, Gevaert SA, Gorog DA, Herrmann J, Lenihan D, Moslehi J, Moura B, Salinger SS, Stephens R, Suter TM, Szmit S, Tamargo J, Thavendiranathan P, Tocchetti CG, van der Meer P, van der Pal HJH, ESC 2022 Scientific Document Group. ESC Guidelines on cardio-oncology developed in collaboration with the European Hematology Association (EHA), the European Society for Therapeutic Radiology and Oncology (ESTRO) and the International Cardio-Oncology Society (IC-OS). Eur Heart J 2022;43:4229–4361. [PubMed: 36017568]

6. Fazal M, Malisa J, Rhee JW, Witteles RM, Rodriguez F. Racial and ethnic disparities in cardio-oncology: a call to Action. JACC CardioOncol 2021;3:201–204. [PubMed: 34308372]

7. Shi T, Jiang C, Zhu C, Wu F, Fotjhadi I, Zarich S. Insurance disparity in cardiovascular mortality among non-elderly cancer survivors. CardioOncology 2021;7:11. [PubMed: 33743837]

8. Dong W, Motairek I, Nasir K, Chen Z, Kim U, Khalifa Y, Freedman D, Griggs S, Rajagopalan S, Al-Kindi SG. Risk factors and geographic disparities in premature cardiovascular mortality in US counties: a machine learning approach. Sci Rep 2023;13:2978. [PubMed: 36808141]

9. Dong W, Bensken WP, Kim U, Rose J, Fan Q, Schiltz NK, Berger NA, Koroukian SM. Variation in and factors associated with US county-level cancer mortality, 2008–2019. JAMA Netw Open 2022;5:e2230925. [PubMed: 36083583]

10. Singh GK. Area deprivation and widening inequalities in US mortality, 1969–1998. Am J Public Health 2003;93:1137–1143. [PubMed: 12835199]

11. Berg KA, Dalton JE, Gunzler DD, Coulton CJ, Freedman DA, Krieger NI, Dawson NV, Perzynski AT. The ADI-3: a revised neighborhood risk index of the social determinants of health over time and place. Health Serv Outcomes Res Method 2021;21:486–509.

12. Gatta G, Botta L, Rossi S, Aareleid T, Bielska-Lasota M, Clavel J, Dimitrova N, Jakab Z, Kaatsch P, Lacour B, Mallone S, Marcos-Gragera R, Minicozzi P, Sánchez-Pérez MJ, Sant M, Santaquilani M, Stiller C, Tavilla A, Trama A, Visser O, Peris-Bonet R, EUROCARE Working Group. Childhood cancer survival in Europe 1999–2007: results of EUROCARE-5-a population-based study. Lancet Oncol 2014;15:35–47. [PubMed: 24314616]

13. Ohman RE, Yang EH, Abel ML. Inequity in cardio-oncology: identifying disparities in cardiotoxicity and links to cardiac and cancer outcomes. J Am Heart Assoc 2021;10:e023852. [PubMed: 34913366]

14. Cousin L, Roper N, Nolan TS. Cardio-oncology health disparities: social determinants of health and care for Black breast cancer survivors. Clin J Oncol Nurs 2021;25:36–41. [PubMed: 34533529]

15. Mackenbach JP, Cavelaars AEJM, Kunst AE, Groenhof F. Socioeconomic inequalities in cardiovascular disease mortality; an international study. Eur Heart J 2000;21:1141–1151. [PubMed: 10924297]

16. Kelly MJ, Weitzen S. The association of lifetime education with the prevalence of myocardial infarction: an analysis of the 2006 behavioral risk factor surveillance system. J Community Health 2010;35:76–80. [PubMed: 19949844]

17. Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. Am J Public Health 1992;82:816–820. [PubMed: 1585961]

18. Rajagopalan S, Landrigan PJ. Pollution and the heart. N Engl J Med 2021;385:1881–1892. [PubMed: 34758254]

19. Motairek I, Sharara J, Makhlouf MHE, Dobre M, Rahman M, Rajagopalan S, Al-Kindi S. Association between particulate matter pollution and CKD mortality by social deprivation. Am J Kidney Dis 2023;81:497–499. [PubMed: 36396086]

20. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020;396:1223–1249. [PubMed: 33069327]

21. Kim DY, Staley F, Curtis G, Buchanan S. Relation between housing age, housing value, and childhood blood lead levels in children in Jefferson County, Ky. Am J Public Health 2002;92:769–772. [PubMed: 11988444]

22. Clark CS, Bornschein RL, Succop P, Que Hee SSQ, Hammond PB, Peace B. Condition and type of housing as an indicator of potential environmental lead exposure and pediatric blood lead levels. Environ Res 1985;38:46–53. [PubMed: 4076111]

23. Rauh VA, Landrigan PJ, Claudio L. Housing and health: intersection of poverty and environmental exposures. Ann N Y Acad Sci 2008;1136:276–288. [PubMed: 18579887]

24. Pun VC, Kazemiparkouhi F, Manjourides J, Suh HH. Long-term PM2.5 exposure and respiratory, cancer, and cardiovascular mortality in older US adults. Am J Epidemiol 2017;186:961–969. [PubMed: 28541385]

25. Rajagopalan S, Al-Kindi SG, Brook RD. Air pollution and cardiovascular disease: JACC State-of-the-Art Review. J Am Coll Cardiol 2018;72:2054–2070. [PubMed: 30336830]

26. Penman-Aguilar A, Carter M, Snead MC, Kourtis AP. Socioeconomic disadvantage as a social determinant of teen childbearing in the U.S. Public Health Rep 2013;128:5–22.

27. Dong W, Bensken WP, Kim U, Rose J, Berger NA, Koroukian SM. Phenotype discovery and geographic disparities of late-stage breast cancer diagnosis across U.S. counties: a machine learning approach. Cancer Epidemiol Biomarkers Prev 2022;31:66–76. [PubMed: 34697059]

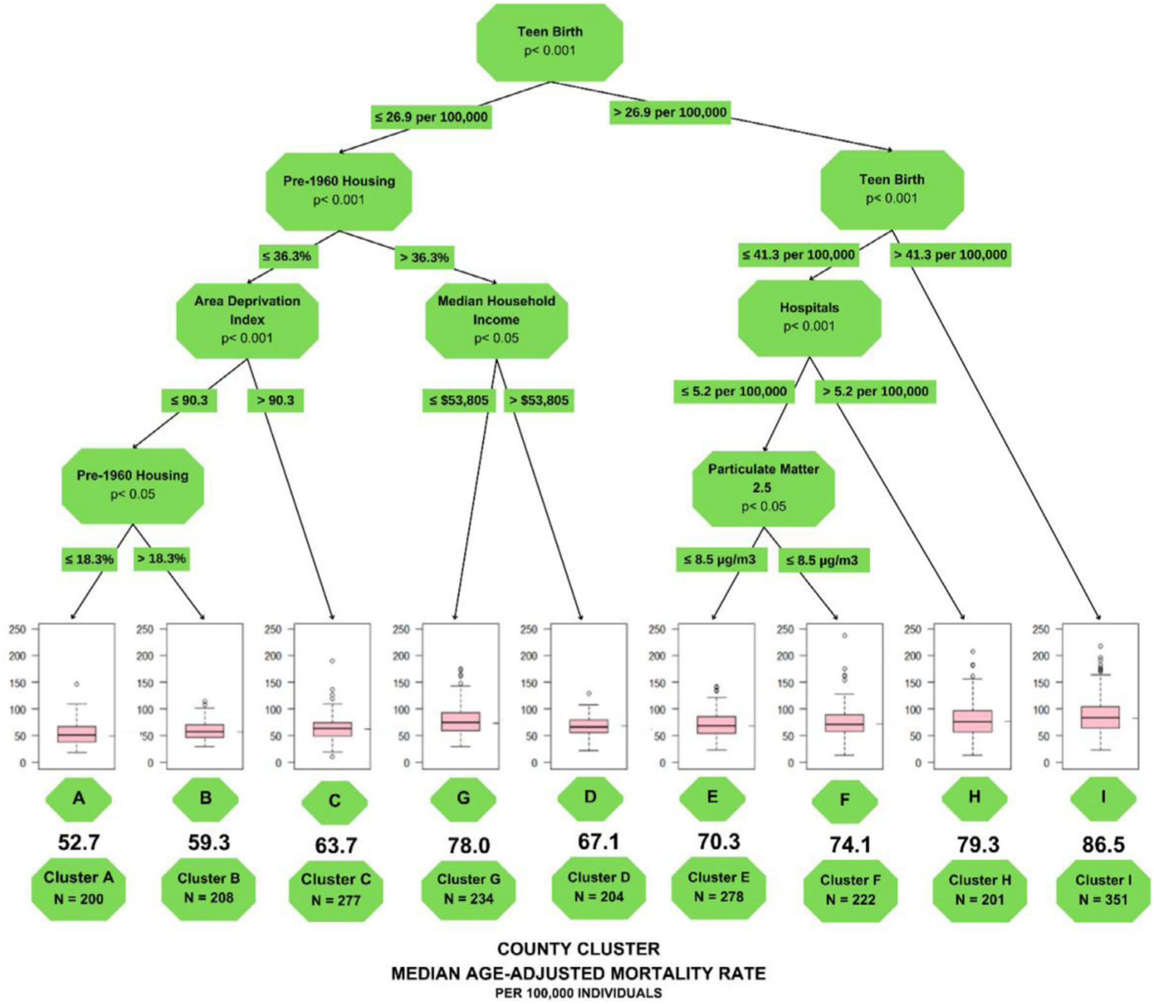

Figure 1.

Classification and Regression Tree analysis to predict county-level COM. Notes: each path down to a terminal node represents a county SEDH cluster. Box plots in the terminal nodes represent age-adjusted COM (per 100,000 individuals). The minimum number of counties in a terminal node was set to 200.

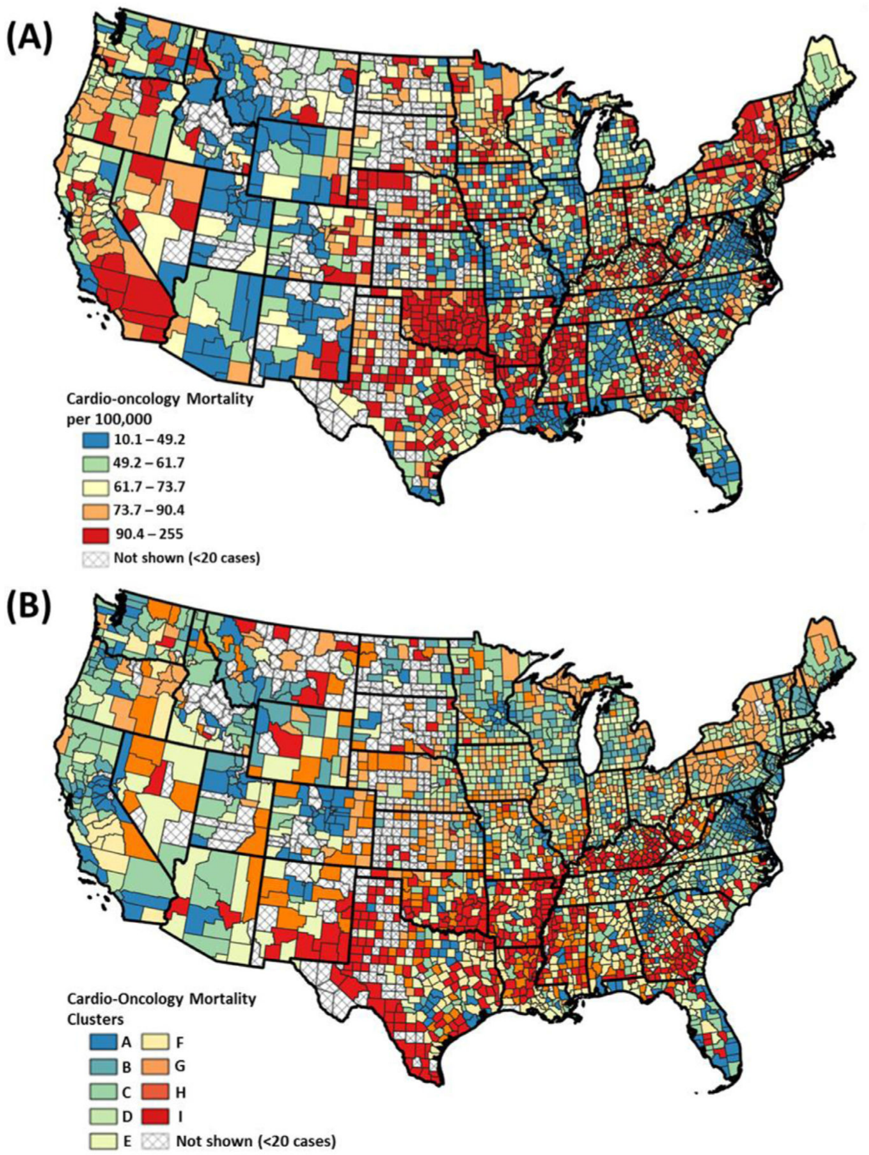

Figure 2.

US County Maps of (*A*) Age-adjusted cardio-oncology mortality (per 100,000 people). (*B*) County cluster of cardio-oncology mortality. Percentages were classified by equal count (quantile) classification method.

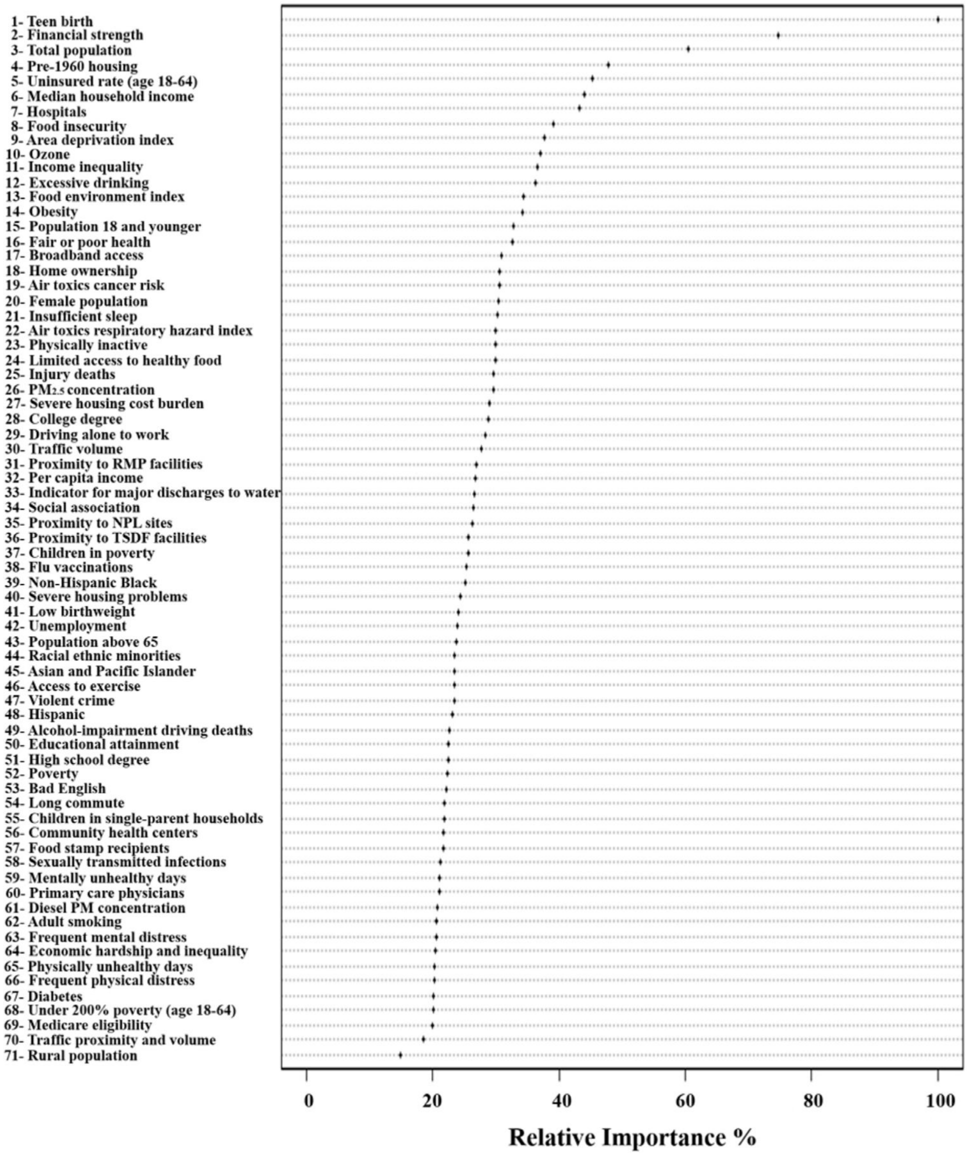

Figure 3.

Dot chart of random forest analysis showing variable importance for predicting county-level age-adjusted cardio-oncology mortality. Notes: the most important variable is at the top and scaled to 100%. The importance of the rest of the variables is shown relative to the top one. NPL = national priorities list; PM = fine particulate matter; RMP = risk management plan; TSDF = Treatment, storage and disposal facilities.

Table 1

Description of variables used in the study’s models

Category	Sources and Variables
Environmental exposure	EPA-EJSCREEN: PM2.5 level in air, air toxics cancer risk, air toxics respiratory hazard index, ozone level in air, diesel PM level in air, traffic proximity and volume, pre-1960 housing (lead paint indicator), proximity to RMP sites, proximity to hazardous waste facilities, proximity to NPL sites, major dischargers to water indicator CHR: Traffic volume
Socioeconomic status	CHR: income inequality, children in poverty, high school degree, college degree, unemployment AHRF: Per capita income, median household income, poverty, under 200% poverty (age 18–64), uninsured rate (age 18–64) R package “sociome”: area deprivation index, financial strength, economic hardship and inequality, and educational attainment
Race/ethnicity	CHR: not proficient in English, Hispanic, Non-Hispanic Black, Asian and Pacific Islander, racial or ethnic minorities
Population age structure	CHR: population age ≤ 18, population age 65+, population female AHRF: Medicare eligibility
Urbanicity	CHR: rural population AHRF: population
Household environment	CHR: children in single-parent households, severe housing problems, severe housing cost burden, homeownership
Health status	CHR: fair or poor health, frequent physical distress, frequent mental distress, physical unhealthy days, mental unhealthy days, insufficient sleep, diabetes, adult obesity, sexually transmitted infections
Health Behavior	CHR: excessive drinking, adult smoking, physical inactivity, violent crime rate
Health outcome	CHR: teen birth, low birthweight, injury death, alcohol-impaired driving deaths, flu vaccinations (Medicare enrollees)
Access to care	AHRF: primary care physicians, hospitals, community health centers
Food access	CHR: food insecurity, limited access to healthy foods, food environment index AHRF: food stamp recipients
Social connectivity	CHR: broadband access, access to exercise opportunities, social associations
Commuting	CHR: driving alone to work, long commute-driving alone

CHR = County Health Rankings & Roadmaps; EPA-EJSCREEN = Environmental Protection Agency - Environmental Justice Screening tool, NLP: national priorities list; PM = fine particulate matter; RMP = risk management plan.

Table 2

County clusters summary statistics and characteristics

County Clusters and Statistics	Characteristics	Cardio-oncology Mortality Ranking[*]
Cluster A N = 200 Median COM rate =52.7 per 100,000 individuals	-Less Teen Births rates -Lowest Pre-1960 Housing -Less socioeconomic vulnerability	9th
Cluster B N = 208 Median COM rate = 59.3 per 100,000 individuals	-Less Teen Births rates -Less Pre-1960 Housing -Less socioeconomic vulnerability	8th
Cluster C N = 277 Median COM rate = 63.7 per 100,000 individuals	-Less Teen Births rates -Less Pre-1960 Housing -More socioeconomic vulnerability	7th
Cluster D N = 204 Median COM rate = 67.1 per 100,000 individuals	-Less Teen Births rates -More Pre-1960 Housing -Higher Median Household Income	6th
Cluster E N = 278 Median COM rate = 70.3 per 100,000 individuals	-More Teen Births rates -Less Hospitals -Less exposure to Particulate Matter 2.5 pollution	5th
Cluster F N = 222 Median COM rate = 74.1 per 100,000 individuals	-More Teen Births rates -Less Hospitals -More exposure to Particulate Matter 2.5 pollution	4th
Cluster G N = 234 Median COM rate = 78.0 per 100,000 individuals	-Less Teen Births rates -More Pre-1960 Housing -Lower Median Household Income	3rd
Cluster H N = 201 Median COM rate = 79.3 per 100,000 individuals	-More Teen Births rates -More Hospitals	2nd
Cluster I N = 351 Median COM rate = 86.5 per 100,000 individuals	-Highest Teen Births rates	1st

COM = cardio-oncology mortality.

[*] Ranking of the county clusters according to median COM mortality rates (from highest to lowest).