Disparities in Cancer Incidence: A Composite Measure to Identify High-Burden Populations, Sites, and Key Drivers in an NCI-Designated Catchment Area

Northwestern Medicine

Robert H. Lurie Comprehensive Cancer Center of Northwestern University

Daniel Antonio, Tarneka M. Manning

Robert H. Lurie Comprehensive Cancer Center of Northwestern University

Background

NCI-Designated Cancer Centers assess cancer burden across diverse populations in their catchment areas, aiming to identify high-burden populations and cancer sites and understand the drivers of disparities. Surveillance data is commonly used to compare cancer outcomes across demographic groups, cancer sites, and geographic regions. However, the vast amount of data and multiple comparisons can make it challenging to interpret trends effectively. We developed a composite disparity measure combining absolute and relative differences to pinpoint cancer sites, counties, and race/ethnicity groups experiencing disproportionate cancer burden in the Robert H. Lurie Comprehensive Cancer Center's catchment area. We also explore associations between social determinants of health and disparities.

Aims

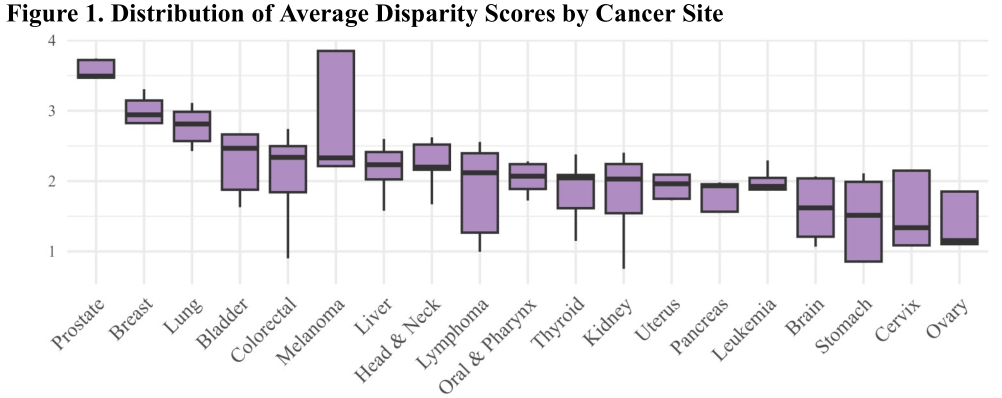
1. Use catchment area surveillance data to develop a composite measure to assess cancer burden disparities among cancer sites, counties, and race/ethnicity groups.
2. Explore the drivers of disparities with the new composite measure by deploying a mixed-effect model controlling for county and cancer site effects.

Methods

Age-adjusted incidence rates (2017-2021) were used to calculate absolute and relative differences for each cancer site and county, comparing race/ethnicity group rates to county averages.(CDC and NCI 2024) Rates below the county average or reported for only one group were excluded. These differences were aggregated by summing across race/ethnicity groups and normalized through scaling and log-transformation. These measures were then averaged to create a composite disparity score. Count-level Social determinants, including labor force participation, insurance coverage, and urban percentage and behavioral risk factors, were obtained from the United States Census Bureau and Center for Disease Control and Prevention(CDC 2023) (United States Census Bureau 2021). Counties were classified as high risk if they had above-average prevalence for three or more factors including, smoking, physical inactivity, obesity, alcohol consumption, and lack of sleep. Mixed-effect linear models, controlling for county and cancer site effects, were utilized to explore risk factors. Absolute and relative disparities were aggregated at the county level and visualized through proportional bar graphs, with composite scores plotted in box plots to show the distribution of disparities across cancer sites.

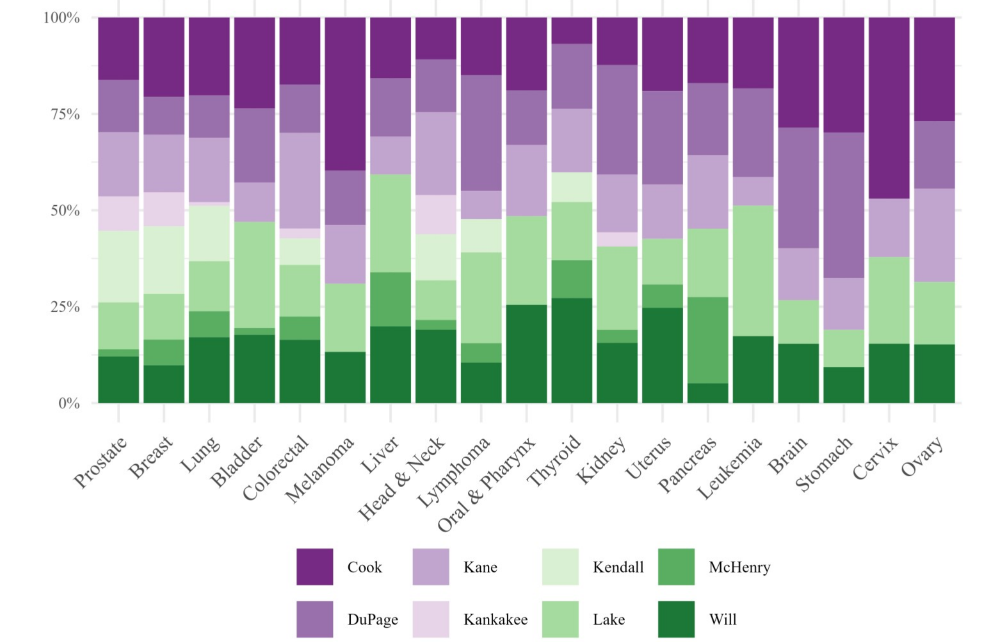
Results

Figure 1. Distribution of Average Disparity Scores by Cancer Site



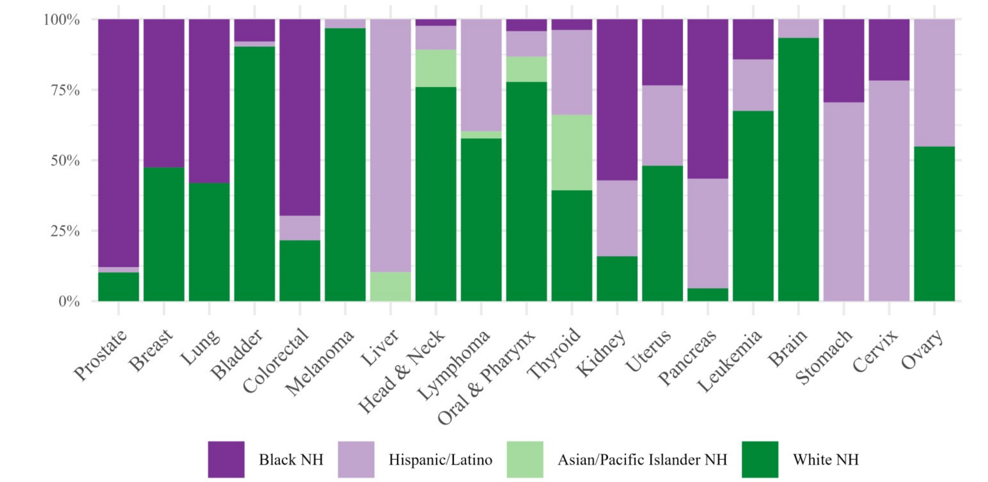
Prostate cancer exhibited the highest average composite disparity score, followed by breast, lung, and bladder cancers. In contrast, stomach, cervical, and ovarian cancers had the lowest scores. With the exception of melanoma, most cancer sites showed similar variability in scores. **Figure 2.** displays the unique contributions of each county to cancer disparities. For instance, Cook County accounts for almost half of the disparity in cervical (47%) and melanoma (40%) cancers, but only a small portion for thyroid (7%) cancer. Kane (19%) and Kendall (17%)

Figure 2. Proportion of Disparity Attributable to Each Catchment Area County



counties contribute the largest proportions of disparities in prostate cancer. DuPage County is responsible for significant disparities in stomach (38%), brain (31%), and lymphoma (30%) cancers. **Figure 3.** illustrates the racial/ethnic disparities across cancer types. Black non-Hispanic individuals account for a large proportion of disparities in prostate (88%) and colorectal (70%) cancers. Hispanic/Latino individuals contribute to a substantial share of disparities in liver (89%), cervical (78%), and stomach (70%) cancers. White non-Hispanic individuals account for nearly all the disparities in melanoma (97%), brain (93%), and bladder

Figure 3. Proportion of Disparity Attributable to Each Race/Ethnicity Group



(90%) cancers. The mixed-effects multivariable model explained much of the variance in the composite disparity score (Conditional $R^2 = 0.80$), **Table 1.**. Urban percentage ($t = 8.89$, $p < 0.001$), insurance coverage ($t = 18.03$, $p < 0.01$), and high-risk factor score ($t = 0.94$, $p < 0.001$) were positively associated with disparities. Recent medical checkups ($t = -79.31$, $p < 0.001$) and labor force participation ($t = -9.42$, $p < 0.001$) were linked to lower scores. The lower variance explained by the social determinants (marginal $R^2 = 0.23$) suggests that county and cancer site variations more strongly drive disparities. The random intercepts indicate that most of the variance in disparities is attributable to site-level differences (variance = 0.34), with county-level variation explaining much less (variance = 0.0002). This is consistent with the random-effects-only model, where cancer site disparities accounted for the largest proportion of variability (variance = 0.33), with county-level factors contributing less (variance = 0.26).

Table 1. Multivariable Liner Mixed Model with Random Effects for County and Cancer Sites

Variable	Estimate	95% CI	P
Annual Labor Force Participation Rate	-9.424	-13, -6.0	<0.001
Insurance Coverage	18.033	9.9, 26	0.003
Urban Percentage	8.888	7.4, 10	<0.001
Recent Checkup	-79.307	-112, -46	<0.001
Behavioral Risk			
Low (ref)			
High	0.937	0.63, 1.2	<0.001
Cancer Site Std. Error	0.58		
County Std. Error	0.014		
Residual Std. Error	0.343		

Conditional $R^2 = 0.800$, marginal $R^2 = 0.23$, No. Obs. = 388; Sigma = 0.343; Log-likelihood = -169; AIC = 355; BIC = 391; REMLcrit = 337; Residual df = 379; County

Conclusion

This study underscores the value of a composite disparity measure in identifying high-disparity populations and their underlying drivers. Prostate, breast and lung, had the highest disparity scores, indicating large differences in incidence rates among these sites. The high burden observed among specific counties and race/ethnicity groups emphasizes the importance of considering geographic, demographic, and site-specific contexts when addressing disparities. The linear mixed models revealed that site-level differences accounted for a larger share of the variance in disparity scores, while county-level factors, influenced by social determinants, also contributed to some variation. Prioritizing high-impact populations and addressing both social determinants of health, regional and site specific factors are all critical for reducing cancer burden disparities and improving health outcomes.

References

CDC. 2023. *PLACES: Local Data for Better Health, County Data 2023 Release*. https://www.cdc.gov/places.

CDC, and NCI. 2024. *National Program of Cancer Registries and Surveillance, Epidemiology, and End Results Program SEER*stat Database: NPCR and SEER Incidence – u.s. Cancer Statistics 2001–2021 Public Use Research Database, 2023 Submission (2001–2021)*. https://www.cdc.gov/united-states-cancer-statistics/public-use/.

United States Census Bureau. 2021. *American Community Survey (2017-2021)*. https://www.census.gov/programs-surveys/acs.